

Krassimir Markov, Vitalii Velychko
(editors)

Applicable Information Models

I T H E A[®]

Sofia

2011

Krassimir Markov, Vitalii Velychko (ed.)

Applicable Information Models

ITHEA®

2011, Sofia, Bulgaria,

ISBN: 978-954-16-0050-4 (printed)

ISBN: 978-954-16-0051-1 (CD)

ISBN: 978-954-16-0052-8 (online)

ITHEA IBS ISC No.: 22

First edition

Recommended for publication by The Scientific Council of the Institute of Information Theories and Applications FOI ITHEA

This issue contains a collection of papers that concern the information modeling and its methodological and technological problems. Main scientific areas which are concerned are: data mining; pattern recognition; decision support systems; robotics; data and knowledge bases; disaster risk assessment, based on heterogeneous information (from satellites and in-situ data, and modeling data); virtual reality in educational systems; telecommunications; open social info-educational platforms; system analysis and synthesis.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

General Sponsor: Consortium FOI Bulgaria (www.foibg.com).

Printed in Bulgaria by ITHEA®

Sofia-1090, P.O.Box 775, Bulgaria

e-mail: info@foibg.com

www.ithea.org

© All rights reserved.

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Copyright © 2011

© 2011 ITHEA® – Publisher; Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org ; e-mail: info@foibg.com

© 2011 Krassimir Markov, Vitalii Velychko – Editors

© 2011 Ina Markova – Technical editor

© 2011 For all authors in the book.

© ITHEA is a registered trade mark.

ISBN: 978-954-16-0050-4 (printed)

ISBN: 978-954-16-0051-1 (CD)

ISBN: 978-954-16-0052-8 (online)

C\o Jusautor, Sofia, 2011

PREFACE

ITHEA International Scientific Society (ITHEA ISS) is aimed to support growing collaboration between scientists from all over the world. The scope of the books of the ITHEA ISS covers the area of Informatics and Computer Science. ITHEA ISS welcomes scientific papers and books connected with any information theory or its application. ITHEA ISS rules for preparing the manuscripts are compulsory.

ITHEA Publishing House is the official publisher of the works of the members of the ITHEA ISS.

Responsibility for papers and books published by ITHEA belongs to authors.

This issue contains a collection of papers that concern the information modeling and its methodological and technological problems. Main scientific areas which are concerned are:

- data mining;
- pattern recognition;
- decision support systems;
- robotics;
- data and knowledge bases;
- disaster risk assessment, based on heterogeneous information (from satellites and in-situ data, and modeling data);
- virtual reality in educational systems;
- telecommunications;
- open social info-educational platforms;
- system analysis and synthesis.

Main part of the articles in this book are written and published in Russian. This way the Russian specialized terminology in the given scientific area became popular and accessible for the ITHEA ISS members who can use Russian language.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

We express our thanks to all authors, reviewers and collaborators as well as to the General Sponsor.

*Sofia – Kiev
June 2011*

K. Markov, V. Velychko

TABLE OF CONTENTS

Preface.....	3
Table of Contents.....	4
Index of Authors.....	6
Построение нелинейных классификаторов в случае многих классов	
<i>Ю.И. Журавлев, Ю. П. Лаптин, А.П. Виноградов.....</i>	<i>7</i>
Кластеризация неполных данных	
<i>Владимир Рязанов, Кирилл Тишин, Антон Щичко.....</i>	<i>14</i>
Групповой многокритериальный выбор вычислительных кластеров	
<i>Алексей Борисович Петровский, Григорий Владимирович Ройзензон.....</i>	<i>23</i>
Прогнозирование динамики пополнения фондов углеводородов	
<i>Михаил Стернин, Геннадий Шепелев.....</i>	<i>30</i>
Решение задачи автоматической сепарации фуллеренсодержащей суспензии	
<i>А.В. Тимофеев, П.Г. Иванов.....</i>	<i>36</i>
Методы и алгоритмы обнаружения объектов для системы управления промышленной безопасностью в рамках концепции создания комплексной АСУ газотранспортной компанией	
<i>Белоус Н.В., Борисенко В.П., Пономарев Ю.В., Борисенко И.В.....</i>	<i>39</i>
Распознавание зданий на спутниковых снимках сверхвысокого разрешения	
<i>Светлана Дьяконова.....</i>	<i>47</i>
Верификация оптико-переменного объекта путём обнаружения изменения цвета	
<i>Екатерина Горшкова, Роман Телятников, Иван Шумский.....</i>	<i>54</i>
Верификация цветных объектов на сложных сценах	
<i>Анатолий Протосавицкий, Екатерина Горшкова, Дмитрий Пилютик, Роман Телятников, Иван Шумский.....</i>	<i>59</i>
Онлайн распознавание рукописных математических формул	
<i>Надеран Эдрис.....</i>	<i>70</i>
Линейное оценивание параметров моделей финансовых временных рядов	
<i>Зайченко Ю.П., Мурга Н.А.....</i>	<i>76</i>
Исследование многокритериальной задачи оптимизации инвестиционного портфеля в нечетких условиях	
<i>Малихех Есфандиярфард , Юрий Зайченко, Ови Нафас Агаи Аг Гамиш.....</i>	<i>83</i>
О свойствах расстояний в вероятностных высказываниях экспертов	
<i>Александр Викентьев, Руслан Викентьев.....</i>	<i>90</i>
Моделирование и оптимизация параметров региональных бюджетов	
<i>Диана Омелянчик.....</i>	<i>97</i>
Применение нейронных сетей в рамках концепции "электронная таможня"	
<i>Борис Мороз, Сергей Коноваленко.....</i>	<i>104</i>

Нейросетевые алгоритмы поиска решения обратной задачи кинематики робота вертикального перемещения (РВП)	
<i>Адиль Тимофеев, Виктор Титов</i>	111
Распределенная информационная система агромониторинга	
<i>Андрей Шелестов, Наталья Морзе, Ольга Куссуль, Юлия Грипич</i>	115
Проектирование базы знаний для систем организационного управления	
<i>Баканова Н.Б., Обознов А.А., Баканов А.С.</i>	125
Модель извлечения знаний из неструктурированных документов корпоративной информационной системы	
<i>Нина Хайрова, Наталья Шаронова</i>	131
Модель виртуального мира мультимедиа тренажера для медицинского образования	
<i>В.В. Грибова, М.В. Петряева, Л.А. Федорищев, М.Ю. Черняховская</i>	140
Модель динамики монетного обращения в средневековой Плиске	
<i>Йордан Табов, Галина Панайотова</i>	149
Формализация модели эволюции телекоммуникационных сетей	155
<i>Галина Гайворонская</i>	155
Особенности применения оптических коммутаторов в современных информационных сетях	
<i>Галина Гайворонская, Александр Рябцов</i>	169
Оценка чувствительности характеристик сетей доступа к вариациям прогнозируемых параметров	
<i>Светлана Сахарова</i>	181
Возможный подход к разработке модели трафика конвергентной телекоммуникационной сети	
<i>Максим Соломицкий</i>	189
Анализ интенсивности нагрузки на реальных телекоммуникационных сетях Юга Украины	
<i>Илья Ганницкий</i>	199
Plagiarism revealing in the source program code on the basis of its systatic representation	
<i>Tetyana Shatovska, Iryna Kamenieva</i>	204
Алгоритм синхронизации объектов распределенной имитационной модели в triiad.Net	
<i>Елена Замятина, Сергей Ермаков</i>	211
Информационно-аналитическая система поддержки управления рисками в региональных производственных комплексах	
<i>Галина Кутергина, Людмила Лядова, Наталья Фролова</i>	221
Учебно-методический комплекс дисциплины «Электрические машины» / Интернет-реализация /	
<i>Евгений Забудский</i>	232
Содержание некоторых этапов системного анализа и синтеза (продолжение «Методологического меморандума» – часть III)	
<i>Анатолий Крислов</i>	244

INDEX OF AUTHORS

<i>Kamenieva I.</i>	204	<i>Мороз Б.</i>	104
<i>Shatovska T.</i>	204	<i>Мурга Н.</i>	76
<i>Баканов А.</i>	125	<i>Обознов А.</i>	125
<i>Баканова Н.</i>	125	<i>Омельянчик Д.</i>	97
<i>Белоус Н.</i>	39	<i>Панайотова Г.</i>	149
<i>Борисенко В.</i>	39	<i>Петровский А.</i>	23
<i>Борисенко И.</i>	39	<i>Петряева М.</i>	140
<i>Викентьев А.</i>	90	<i>Пилюттик Д.</i>	59
<i>Викентьев Р.</i>	90	<i>Пономарев Ю.</i>	39
<i>Виноградов А.</i>	7	<i>Протосавицкий А.</i>	59
<i>Гайворонская Г.</i>	155, 169	<i>Ройзензон Г.</i>	23
<i>Гамиш О.Н.А.А.</i>	83	<i>Рябцов А.</i>	169
<i>Ганницкий И.</i>	199	<i>Рязанов В.</i>	14
<i>Горшкова Е.</i>	54, 59	<i>Сахарова С.</i>	181
<i>Грибова В.</i>	140	<i>Соломицкий М.</i>	189
<i>Грипич Ю.</i>	115	<i>Стернин М.</i>	30
<i>Дьяконова С.</i>	47	<i>Табов Й.</i>	149
<i>Ермаков С.</i>	211	<i>Телятников Р.</i>	54, 59
<i>Есфандиярфард М.</i>	83	<i>Тимофеев А.</i>	36, 111
<i>Журавлев Ю.</i>	7	<i>Титов В.</i>	111
<i>Забудский Е.</i>	232	<i>Тишин К.</i>	14
<i>Зайченко Ю.</i>	76, 83	<i>Федорищев Л.</i>	140
<i>Замятина Е.</i>	211	<i>Фролова Н.</i>	221
<i>Иванов П.</i>	36	<i>Хайрова Н.</i>	131
<i>Коноваленко С.</i>	104	<i>Черняховская М.</i>	140
<i>Крисипов А.</i>	244	<i>Шаронова Н.</i>	131
<i>Куссуль О.</i>	115	<i>Шелестов А.</i>	115
<i>Кутергина Г.</i>	221	<i>Шепелев Г.</i>	30
<i>Лаптин Ю.</i>	7	<i>Шумский И.</i>	54, 59
<i>Лядова Л.</i>	221	<i>Щичко А.</i>	14
<i>Морзе Н.</i>	115	<i>Эдрис Н.</i>	70

ПОСТРОЕНИЕ НЕЛИНЕЙНЫХ КЛАССИФИКАТОРОВ В СЛУЧАЕ МНОГИХ КЛАССОВ

Ю.И. Журавлев, Ю. П. Лаптин, А.П. Виноградов

Abstract: Рассматриваются задачи построения нелинейных классификаторов для случая многих классов. Для произвольной обучающей выборки формулируется NP-трудная задача построения классификатора заданной сложности, обеспечивающего минимизацию эмпирического риска. Рассмотрена непрерывная релаксация задачи минимизации эмпирического риска, предложены некоторые новые подходы к решению таких задач

Keywords: кластер, решающее правило, дискриминантная функция, линейное и нелинейное программирование, негладкая оптимизация

ACM Classification Keywords: G.1.6 Оптимизация – градиентные методы, I.5 Распознавание образов; I.5.2 Методология проектирования – построение и оценивание классификаторов

Acknowledgement: Работа поддержана совместным грантом НАН Украины и РФФИ № 10-01-90419

Введение

Проблемы построения классификаторов, линейных и нелинейных, обычно рассматриваются для двух непересекающихся множеств. Для линейно разделимых множеств естественным образом формулируются и эффективно решаются задачи построения оптимальных классификаторов [Мест, 2008], [В, 2009], [ШГ, 2004], [Маз, 2010], [DN, 2009]. Для линейно неразделимых множеств, как формулировка, так и решение задач построения классификаторов наталкиваются на определенные трудности. При построении линейных классификаторов в этом случае часто используются эвристические алгоритмы [В, 2009], [ВМ, 1992], [СБЖК, 2009]. Широко используется метод опорных векторов, требующий задания значений некоторых параметров. Выбор этих значений далеко не очевиден. Альтернативой для линейно неразделимых множеств является переход к нелинейным классификаторам – квадратичным, полиномиальным. По-видимому, увеличивая сложность классификатора, всегда можно добиться правильного разделения исходных множеств. Такая цель не всегда оправдана, поскольку исходные множества (обучающие выборки) могут формироваться в условиях возможного возникновения тех, или других ошибок. Вопросы постановки и решения задач построения классификаторов в условиях неопределенности рассматриваются в ([BTGN, 2009], Chapter 12). Естественным в этой ситуации является построение классификатора заданного типа (заданной сложности), минимизирующего число точек обучающей выборки, разделяемых неправильно (минимизация эмпирического риска).

Замечательной особенностью рассматриваемых задач является то, что для достаточно широкого класса нелинейных классификаторов условие правильного разделения множеств классификатором эквивалентно тому, что параметры этого классификатора удовлетворяют некоторой системе линейных неравенств [Маз, 2010]. Эта особенность позволяет создавать эффективные алгоритмы формирования таких нелинейных классификаторов.

В настоящей работе задачи построения нелинейных классификаторов рассматриваются для случая многих классов. Эта работа является развитием результатов, полученных ранее для линейных классификаторов [LLV, 2010], [ZhLV, 2010].

1. Постановка задачи

Для заданной совокупности конечных непересекающихся множеств $\Omega_i \subset R^n$, $i=1, \dots, m$ (обучающей выборки из m непересекающихся классов) рассматривается задача построения алгоритма классификации следующего вида

$$a(x, W) = \arg \max_i \left\{ f_i(x, W^i) : i = 1, \dots, m \right\}, \quad x \in R^n, \quad W^i \in R^{L_i+1} \quad (1)$$

где W^i – совокупность параметров функции $f_i(x, W^i)$, $i = 1, \dots, m$, $W = (W^1, \dots, W^m)$. Функции $f_i(x, W^i)$ называются (приоритетными) весовыми функциями.

Говорят, что классификатор $a(x, W)$ правильно разделяет точки из Ω_i , $i=1, \dots, m$, если $a(x, W) = i$, для всех $x \in \Omega_i$, $i=1, \dots, m$. Если функции $f_i(x, W^i)$ линейны – $f_i(x, W^i) = (w^i, x) + w_0^i$, $i=1, \dots, m$, то классификатор $a(x, W)$ называется линейным. Задачи построения линейных классификаторов рассматривались в [LLV, 2010], [ZhLV, 2010].

Будем предполагать, что для каждого класса i задана совокупность базисных функций $\phi_0^i(x) \equiv 1$, $\phi_j^i(x)$, $j = 1, \dots, L_i$, на основе которых формируются весовые функции $f_i(x, W^i)$, $W^i = (w_0^i, w_1^i, \dots, w_{L_i}^i)$, $i = 1, \dots, m$

$$f_i(x, W^i) = \sum_{j=1}^{L_i} w_j^i \phi_j^i(x) + w_0^i \quad (2)$$

Совокупности базисных функций могут совпадать для разных классов, например, $\phi_j^i(x) = x_j$, $j = 1, \dots, n$, $i = 1, \dots, m$ для линейных классификаторов. В общем случае совокупность базисных функций для каждого класса может быть своя.

В зависимости от выбора совокупности базисных функций рассматриваются задачи построения квадратичных, полиномиальных или других алгоритмов классификации.

Множества Ω_i , $i = 1, \dots, m$ называются разделимыми в заданной совокупности базисных функций, если существует классификатор, сформированный на основе этой совокупности базисных функций, правильно разделяющий точки из этих множеств.

Обозначим $\Omega = \bigcup_{i=1}^m \Omega_i$. Пусть точки множества Ω перенумерованы, T – совокупность индексов,

$$\Omega = \{x^t : t \in T\}, \quad T_i - \text{совокупность индексов множества } \Omega_i, \quad \Omega_i = \{x^t : t \in T_i\}, \quad T = \bigcup_{i=1}^m T_i.$$

Положим $i(t)$ – номер множества Ω_i , которому принадлежит точка x^t , $t \in T$. Величина

$$g^t(W) = \min \left\{ f_i(x^t, W^i) - f_j(x^t, W^j) : j \in \{1, \dots, m\} \setminus i, i = i(t) \right\} \quad (3)$$

называется отступом (margin) или зазором (gap) классификатора $a(x, W)$ на точке x^t , $t \in T$.

Обозначим $z_j^{it} = \phi_j^i(x^t)$, тогда зазор $g^t(W)$ можно представить в виде

$$g^t(W) = \min \left\{ \sum_{k=1}^{L_i} w_k^i z_k^{it} - \sum_{k=1}^{L_j} w_k^j z_k^{jt} + w_0^i - w_0^j : j \in \{1, \dots, m\} \setminus i, i = i(t) \right\} \quad (4)$$

Заметим, что $g^t(W)$ – вогнутая кусочно-линейная функция. Классификатор $a(x, W)$ допускает ошибку на точке x^t тогда и только тогда, когда зазор $g^t(W)$ отрицателен.

Величина $g(W) = \min \{g^t(W) : t \in T\}$ называется зазором классификатора $a(x, W)$ на совокупности множеств $\Omega_i, i = 1, \dots, m$.

Классификатор $a(x, W)$ правильно разделяет точки из $\Omega_i, i = 1, \dots, m$, если $g(W) > 0$.

Замечание 1. Классификатор $a(x, W)$ инвариантен относительно умножения всех функций f_i на положительное число, зазор $g(W)$ линеен относительно такой операции умножения. Классификатор $a(x, W)$ и отступ $g(W)$ инвариантны относительно добавления произвольного числа ко всем f_i .

Величину $g(W)$ можно использовать как критерий качества классификатора $a(x, W)$ (чем больше значение $g(W)$, тем надёжнее разделяются точки из $\Omega_i, i = 1, \dots, m$), однако, при этом должна учитываться некоторая нормировка совокупности векторов W , которую обозначим $\eta(W)$ и будем называть нормой классификатора $a(x, W)$. Вопрос выбора нормировки имеет самостоятельное значение и должен рассматриваться отдельно.

В дальнейшем будем использовать функцию:

$$\eta(W) = \sqrt{\sum_{i=1}^m \sum_{j=1}^{L_i} (w_j^i)^2} \quad (5)$$

В качестве нормы $\eta(W)$ могут использоваться также другие функции [LLV, 2010]. Норма вида (5) выбрана из тех соображений, чтобы для случая линейных классификаторов и двух множеств (классов) Ω_1 и Ω_2 сформулированные задачи были эквивалентны традиционно рассматриваемым, в частности, задаче, которая решается в методе опорных векторов.

С учетом введенных обозначений задачу построения оптимального классификатора (определения значений параметров W) запишем в следующем виде: найти

$$g^* = \max_W \{g(W) : \eta(W) \leq 1, W \in R^L\} \quad (6)$$

где L – размерность вектора W . Поскольку вектор $W = 0$ является допустимым, то задача (6) имеет решение всегда и $g^* \geq g(0) = 0$. Заметим, что $g^* > 0$, если множества $\Omega_i, i = 1, \dots, m$ разделимы в заданной совокупности базисных функций, т.е. существует классификатор, правильно разделяющий эти множества. Рассмотрим также задачу: найти

$$\eta^* = \min_V \{\eta(V) : g(V) \geq 1, V \in R^L\} \quad (7)$$

Близкие по смыслу задачи рассматривались разными авторами (см., напр., [4, 9]).

Лемма 1 [LLV, 2010]. Пусть W^* – оптимальное решение задачи (6). Тогда если $g^* > 0$, то задача (7) также имеет оптимальное решение V^* и $V^* = W^* / g^*$, $\eta^* = 1/g^*$, если $g^* = 0$, то задача (7) не имеет допустимых решений.

Рассмотрим более подробно задачи построения нелинейных классификаторов. Учитывая (4), представим задачу (6) в форме задачи линейного программирования с дополнительным квадратичным ограничением: найти

$$g^* = \max_{w, \delta} \delta \quad (8)$$

при ограничениях

$$\sum_{k=1}^{L_i} w_k^i z_k^{it} - \sum_{k=1}^{L_j} w_k^j z_k^{jt} + w_0^i - w_0^j \geq \delta, \quad j \in \{1, \dots, m\} \setminus i, t \in T_i, i = 1, \dots, m \quad (9)$$

$$\sum_{i=1}^m \sum_{j=1}^{L_i} (w_j^i)^2 \leq 1 \quad (10)$$

Соответственно, задача (7) есть задача квадратичного программирования: найти

$$\eta^* = \min_v \sum_{i=1}^m \sum_{j=1}^{L_i} (v_j^i)^2 \quad (11)$$

при ограничениях

$$\sum_{k=1}^{L_i} v_k^i z_k^{it} - \sum_{k=1}^{L_j} v_k^j z_k^{jt} + w_0^i - w_0^j \geq 1, \quad j \in \{1, \dots, m\} \setminus i, t \in T_i, i = 1, \dots, m \quad (12)$$

Можно показать, что в случае линейных классификаторов и $m = 2$ задача (11)-(12) эквивалентна задаче, которая решается с целью построения полосы максимальной ширины, разделяющей линейно разделимые множества Ω_1, Ω_2 .

Для решения рассмотренных задач может использоваться существующее эффективное программное задач оптимизации общего назначения, если число точек в обучающей выборке не очень велико. Результаты такого рода вычислительных экспериментов приведены в [LLV, 2010]. При большом числе точек в обучающей выборке целесообразно использовать методы негладкой оптимизации [СБЖК, 2009], [Sh, 1998].

Задачи (8)-(10) и (11)-(12) позволяют находить оптимальный классификатор только в случае, когда множества $\Omega_i, i = 1, \dots, m$ разделимы в заданной совокупности базисных функций. В противном случае совокупность базисных функций можно расширить, т.е. от линейных классификаторов можно перейти к квадратичным или к полиномиальным более высокого порядка. Такие переходы к полиномиальным классификаторам должны быть обоснованы содержательно в результате анализа рассматриваемой прикладной задачи, поскольку неразделимость множества $\Omega_i, i = 1, \dots, m$ в заданной совокупности базисных функций может быть связана не со сложностью конфигурации классов, а с ошибками определения координат точек множеств Ω_i .

Таким образом, возникает необходимость разработки таких формулировок задач, которые позволяют строить рациональные классификаторы в случае, когда множества Ω_i неразделимы в заданной совокупности базисных функций.

Для случая двух множеств Ω_1, Ω_2 в ряде работ (см., напр. [В, 2009], [ВМ, 1992], [СБЖК, 2009]) предлагаются эвристические формулировки задач, позволяющие находить линейные классификаторы

для линейно неразделимых множеств. Наиболее известным в этой связи является метод опорных векторов (см., напр. [В, 2009]).

2. Минимизация эмпирического риска

В случае, когда множества $\Omega_i, i = 1, \dots, m$, неразделимы в заданной совокупности базисных функций, естественным критерием выбора классификатора является минимизация эмпирического риска, т.е. числа точек обучающей выборки, которые классификатор разделяет неправильно.

Будем считать, что задан некоторый параметр $\bar{\delta} > 0$ надежности разделения точек обучающей выборки $\Omega_i, i = 1, \dots, m$. Будем говорить, что точки $x^t, t \in T$, для которых величина зазора $g^t(W) < \bar{\delta}$, разделяются классификатором $a(x, W)$ ненадежно.

В дальнейшем значение эмпирического риска будем определять с учетом надежности, определяемой параметром $\bar{\delta}$ – эмпирический риск равен числу точек обучающей выборки, которые классификатор разделяет неправильно или ненадежно.

Лемма 3. Пусть $x^\alpha \in \Omega_i, x^\beta \in \Omega_j$ классификатор $a(x, W)$ правильно разделяет эти точки, для нормы классификатора выполняется ограничение (10). Тогда существует конечное число $R > 0$ такое, что

$$-R \leq w_0^i - w_0^j \leq R \quad (13)$$

Доказательство. Для точек x^α, x^β можно записать неравенства

$$\sum_{k=1}^{L_j} w_k^j z_k^{j\alpha} - \sum_{k=1}^{L_i} w_k^i z_k^{i\alpha} \leq w_0^i - w_0^j \leq \sum_{k=1}^{L_j} w_k^j z_k^{j\beta} - \sum_{k=1}^{L_i} w_k^i z_k^{i\beta}. \text{ Учитывая (10), можно показать, что}$$

$$-\sqrt{\sum_{k=1}^{L_j} (z_k^{j\alpha})^2} - \sqrt{\sum_{k=1}^{L_i} (z_k^{i\alpha})^2} \leq w_0^i - w_0^j \leq \sqrt{\sum_{k=1}^{L_j} (z_k^{j\beta})^2} + \sqrt{\sum_{k=1}^{L_i} (z_k^{i\beta})^2}. \text{ Откуда следует утверждение леммы!}$$

Каждой точке $x^t, t \in T$ поставим в соответствие переменную $y_t = 0 \vee 1$ так, что $y_t = 0$, если точка x^t учитывается при формировании задачи (8)-(10), $y_t = 1$ – в противном случае. Пусть задано большое положительное число B . Задача минимизации эмпирического риска имеет вид: найти

$$Q^* = \min_{w, y} \left\{ \sum_{t \in T} y_t \right\} \quad (14)$$

при ограничениях

$$\sum_{k=1}^{L_j} w_k^j z_k^{jt} - \sum_{k=1}^{L_i} w_k^i z_k^{it} + w_0^i - w_0^j \geq \bar{\delta} - B \cdot y_t, \quad j \in \{1, \dots, m\} \setminus i, \quad t \in T_i, \quad i = 1, \dots, m \quad (15)$$

$$\eta(W) \leq 1 \quad (16)$$

$$\sum_{t \in T_i} y_t \leq |T_i| - 1, \quad i = 1, \dots, m \quad (17)$$

$$0 \leq y_t \leq 1, \quad t \in T \quad (18)$$

$$y_t = 0 \vee 1, \quad t \in T \quad (19)$$

Из ограничений (16), (13) следует, что если $y_t = 1$, то при достаточно большом значении числа B соответствующие неравенства вида (15) выполняются всегда, т.е. точка x^t исключается из задачи.

Ограничения (17) определяют условие того, что, по крайней мере, одна точка из каждого множества Ω_i должна быть включена в задачу. Оптимальное значение Q^* равно минимальному эмпирическому риску с учетом надежности $\bar{\delta}$. Задача (14)-(19) является NP -трудной, но приближенные решения могут быть найдены сравнительно просто по решению непрерывной релаксации этой задачи и могут быть уточнены в рамках общей схемы метода ветвей и границ. При формулировке задачи минимизации эмпирического риска могут учитываться дополнительные условия и информация.

Для вычисления оценки снизу для величины Q^* (минимального эмпирического риска) рассмотрим непрерывную релаксацию сформулированной задачи – задачу (14)-(18). Оптимальное значение релаксированной задачи обозначим q^* . Для ее решения будем использовать схему декомпозиции по переменным W . Пусть переменные W зафиксированы. Учитывая (4), задачу минимизации по переменным y представим в следующем виде: найти

$$q(W) = \min_y \left\{ \sum_{t \in T} y_t \right\} \quad (20)$$

при ограничениях

$$y_t \geq \frac{1}{B} (\bar{\delta} - g^t(W)), t \in T \quad (21)$$

$$\eta(W) \leq 1 \quad (22)$$

$$\sum_{t \in T_i} y_t \leq |T_i| - 1, i = 1, \dots, m \quad (23)$$

$$0 \leq y_t \leq 1, t \in T \quad (24)$$

Обозначим $d^t(W) = \max \left(0, \frac{1}{B} (\bar{\delta} - g^t(W)) \right)$. Очевидно, что если задача (20)-(24) имеет решение, то

$y^t = d^t(W)$. Откуда получаем задачу минимизации по переменным W : найти

$$q^* = \min \sum_{t \in T} d^t(W) \quad (25)$$

при ограничениях

$$\eta(W) \leq 1 \quad (26)$$

$$\sum_{t \in T_i} d^t(W) \leq |T_i| - 1, i = 1, \dots, m \quad (27)$$

$$d^t(W) \leq 1, t \in T \quad (28)$$

Функции $d^t(W)$ – выпуклые кусочно-линейные, $\eta(W)$ – квадратичная положительно определенная.

Задача (25)-(28) для случая линейных классификаторов рассматривалась в [ZhLV, 2010]. Было показано, что лагранжева релаксация этой задачи при специальном подборе значений множителей Лагранжа эквивалентна задаче, которая решается в методе опорных векторов.

Для решения задачи (25)-(28) вычисления нижней оценки минимального эмпирического риска может использоваться метод негладких штрафных функций и эффективные алгоритмы безусловной выпуклой оптимизации [Sh, 1998]. При использовании такого подхода требуется каким-либо образом подбирать значения штрафных коэффициентов.

Другими подходами являются использование выпуклых конических продолжений [ЛЛ, 2010] или конических аппроксимаций целевой функции [ЛБ, 2011].

Заключение

В работе для случая многих классов рассмотрены задачи построения нелинейных классификаторов, весовые функции которых представимы в виде линейной комбинации заданной совокупности базисных (в общем случае нелинейных) функций. При условии того, что обучающая выборка разделима в заданной совокупности классификаторов, естественным образом формулируются и эффективно решаются задачи построения оптимальных классификаторов. В случае неразделимости обучающей выборки формулируется задача построения классификатора, обеспечивающего минимизацию эмпирического риска. Такая задача является частично целочисленной и NP-трудной. Для практического использования рассмотрена непрерывная релаксация задачи минимизации эмпирического риска, которая сводится к задаче выпуклого программирования с кусочно-линейными функциями, и предложены некоторые новые подходы к решению таких задач.

References

- [BM, 1992] Bennett K.P., Mangasarian O.L. Robust Linear Programming Discrimination of Two Linearly Inseparable Sets // Optimization Methods and Software. – 1992. – №5. – P. 23-34.
- [BTGN, 2009] Ben-Tal A., Ghaoui L.E., Nemirovski A. Robust Optimization. – Princeton University Press, 2009. – 542 p.
- [DD, 2009] E. Dogantekin, A. Dogantekin, D. Avci Automatic Hepatitis Diagnosis System based on Linear Discriminant Analysis and Adaptive Network Based Fuzzy Inference System // Expert Systems with Applications, In Press, 2009.
- [DN, 2009] Koel Das, Zoran Nenadic. An efficient discriminant-based solution for small sample size problem // Pattern Recognition – Volume 42, Issue 5, 2009, Pages 857-866.
- [LLV, 2010] Laptin Yu. P., Likhovid A. P., and Vinogradov A. P. Approaches to Construction of Linear Classifiers in the Case of Many Classes // Pattern Recognition and Image Analysis, Vol. 20, No. 2, 2010, p. 137-145.
- [Sh, 1998] Shor N.Z. Nondifferentiable Optimization and Polynomial Problems. – Dordrecht, Kluwer, 1998. – 394 p.
- [ZhLV, 2010] Zhuravlev Yu., Laptin Yu, Vinogradov A. Minimization of empirical risk in linear classifier problem // New Trends in Classification and Data Mining, ITHEA, Sofia, Bulgaria, 2010. – Pages 9-15
- [B, 2009] Воронцов К.В. Машинное обучение. – [http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_\(курс_лекций%2C_К.В.Воронцов\)](http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_(курс_лекций%2C_К.В.Воронцов)) – Последнее изменение: 30 мая 2009
- [ГС, 2008] Гупал А.М., Сергиенко И.В. Оптимальные процедуры распознавания. - Киев: Наук. думка, 2008. - 232 с.
- [ЛБ, 2011] Лаптин Ю.П., Бардадым Т.А. Некоторые подходы к регуляризации нелинейных задач оптимизации // Проблемы управления и информатики. 2011, № 3. – С. 57-68.
- [ЛЛ, 2010] Лаптин Ю.П. Лиховид А.П. Использование выпуклых продолжений функций для решения нелинейных задач оптимизации // Управляющие системы и машины. 2010, № 6. – С. 25–31.
- [Маз, 2010] Мазуров Вл.Д. Математические методы распознавания образов.– Екатеринбург, 2010. – 101 с. – Уральский государственный университет им. А.М.Горького.
- [Мест, 2008] Местецкий Л.М. Математические методы распознавания образов. – <http://www.intuit.ru/department/graphics/imageproc/> – Опубликовано 30.04.2008
- [СБЖК, 2009] Методи негладкої оптимізації у спеціальних задачах класифікації, Стецюк П.І., Березовський О.А., Журбенко М.Г., Кропотов Д.О. – Київ, 2009. – 28 с. – (Препр. НАН України. Ін-т кібернетики ім. В.М.Глушкова; 2009–1)
- [ШГ, 2004] Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию. – К.: Наукова думка, 2004. – 545 с.

Информация об авторах

Ю.И. Журавлев – академик РАН, заместитель директора по научной работе Вычислительного центра им. А. А. Дородницына РАН, ул. Вавилова 40, 119333 Москва, Российская федерация; e-mail: zhuravlev@ccas.ru

Ю.П. Лаптин – старший научный сотрудник, Институт кибернетики им. В.М. Глушкова НАН Украины, пр. академика Глушкова, 40, 03650 Киев, Украина; e-mail: laptin_yu_p@mail.ru

А.П. Виноградов – старший научный сотрудник, Вычислительный центр им. А. А. Дородницына РАН, ул. Вавилова 40, 119333 Москва, Российская федерация; e-mail: vngrccas@mail.ru

КЛАСТЕРИЗАЦИЯ НЕПОЛНЫХ ДАННЫХ

Владимир Рязанов, Кирилл Тишин, Антон Щичко

Abstract: В работе рассматриваются различные подходы к решению задач кластеризации данных при наличии в них пропусков. Применяются методы коллективных решений, а также алгоритмы заполнения пропущенных значений признаков. Также предлагаются способы вычисления оценок качества кластеризации данных с прочерками.

Keywords: кластеризация, неполные данные, прочерки, коллективные решения

ACM Classification Keywords: I.5.3 Computing Methodologies - Pattern recognition - Clustering

Введение

Случаи наличия пропусков в обучающих выборках являются обычным явлением в практической классификации без учителя (кластеризации). В настоящее время сформировалось два общих подхода для решения задач кластеризации данных с пропусками. Первый подход состоит в восстановлении пропущенных значений признаков и последующей кластеризации полученных полных данных. При втором подходе осуществляется прямая кластеризация неполных данных. Оба подхода имеют свои преимущества и недостатки. При первом подходе требуется создание методов восстановления значений признаков, при их восстановлении теряется некоторая информация. Однако здесь возможно последующее использование существующих алгоритмов и программ кластеризации полных данных. При прямой кластеризации неполных данных мы не теряем дополнительно информацию, но здесь необходимо создание новых алгоритмов кластеризации неполных данных или модификация известных методов на случай неполноты данных. Существуют многочисленные алгоритмы для восстановления значений признаков и ряд методов прямой кластеризации неполных данных.

Алгоритмы восстановления значений признаков можно условно разделить на два типа. Первый тип алгоритмов (marginalization) предполагает простое исключение из обучающих выборок неполных признаковых описаний. Ясно, что этот путь может быть целесообразным только при относительно малой доле неполных объектов. При втором типе алгоритмов пропущенные значения признаков заменяются их оценками (imputation). Здесь применяются обычно простейшие подходы (замена прочерков на усредненные статистические оценки значений признаков: means, random, the nearest neighbor method, и т.п.) [1,2], и регрессионные модели, когда неизвестное значение признака вычисляется с помощью найденной функции регрессии по известным признакам (линейная регрессия, SVR [3]). Широкое распространение получил EM-алгоритм, предполагающий вероятностную модель построения выборок [1].

Второй подход при кластеризации неполных данных состоит в адаптации методов к случаям неполных данных. В данном случае не требуется предварительного восстановления значений признаков. В [4] предложена модификация метода «fuzzy k-means». В работе [5] описаны два метода разбиения неполных данных на линейные нечеткие кластеры. Модификация метода «k-means», где из-за пропусков создаются системы ограничений, изложена в [6].

Результаты сравнения различных подходов показывают, что при решении практических задач априори не ясно, какой из подходов окажется наиболее приемлемым. Таким образом, создание новых подходов и алгоритмов кластеризации неполных данных является актуальной задачей. Другой актуальной проблемой в данной области является оценка качества кластеризаций (степени «определенности» кластеризаций) как результат отсутствия значений некоторых признаков.

В настоящей статье предлагается два подхода для решения задач кластеризации с прочерками на заданное число кластеров с оценкой степени определенности решений. Первый подход основан на восстановлении значений неизвестных признаков, решении задачи кластеризации данных без прочерков и вычислении степени определенности полученных кластеризаций на основе оценки их устойчивости. Второй подход основан на решении конечного множества задач кластеризации $Z_i, i = 1, \dots, N$, выборки допустимых полных описаний, соответствующих выборкам исходных неполных описаний, и построении коллективного решения. Разброс решений задач $Z_i, i = 1, \dots, N$, относительно коллективного решения и используется для оценки степени определенности кластеризации неполных данных (коллективного решения).

Кластеризация неполных данных на базе восстановления значений признаков

Пусть дана стандартная выборка неполных признаков описаний объектов $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$. Считаем, что $x_{ij} = \Delta, \forall j \in \Omega_i, i = 1, \dots, m$, где символом Δ обозначаем неизвестное значение признака (пропуск, прочерк). Иногда также будем использовать обозначение $J = \{\langle i, j \rangle, i = 1, 2, \dots, m, j \in \Omega_i\}$. Будем использовать локальный метод заполнения прочерков, суть которого состоит в следующей итерационной процедуре. Сначала все неизвестные значения заполняются случайными числами из области допустимых значений признака $x_{ij} \in M_j \subseteq R, j = 1, 2, \dots, n$, (область M_j определяется как конечное множество значений j -го признака, которые принимают объекты обучающей выборки). Далее неизвестные значения последовательно модифицируются с помощью сочетания метода k -ближайших соседей и процедуры сдвига.

Пусть фиксирована метрика ρ в R и значение целочисленного параметра $k, 1 \leq k \leq m - 1$.

Локальный алгоритм восстановления значений признаков [7]:

Шаг 0. Инициализация случайных $x_{ij}^0 \in M_j, j = 1, 2, \dots, n$, если $x_{ij} = \Delta, \forall j \in \Omega_i, i = 1, \dots, m$. Получаем таблицу полных описаний $\|x_{ij}^{(0)}\|_{m \times n}$.

Шаг $t=1, 2, \dots$. Пусть $x_{ij}^{(t-1)*}$ - среднее значение признака № j по k ближайшим соседям объекта \mathbf{x}_i .

Тогда определяем $x_{ij}^{(t)} = x_{ij}^{(t-1)} + \theta(x_{ij}^{(t-1)*} - x_{ij}^{(t-1)}), \forall \langle i, j \rangle \in J$. Здесь $0 < \theta \leq 1$ управляющий параметр (скорость обучения). Если не выполняется условие останова, шаг повторяется.

В качестве критерия останова используем один из стандартных критериев или их комбинацию (максимальное число итераций $N, \rho(x_{ij}^{(t)}, x_{ij}^{(t-1)}) \leq \varepsilon, \forall \langle i, j \rangle \in J$, и др.). После вычисления значений

$x_{ij}^{(t)}, \forall \langle i, j \rangle \in J$ положим x_{ij}^* равным значению из M_j , ближайшему к соответствующему $x_{ij}^{(t)}, \forall \langle i, j \rangle \in J$. Если $x_{ij} \neq \Delta$, положим $x_{ij}^* = x_{ij}$.

Полученную в итоге выборку полных описаний обозначим как $X^* = \{\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_m^*\}$. Пусть получено решение задачи кластеризации данной выборки на l кластеров некоторым алгоритмом A :

$K = \{K_1, K_2, \dots, K_l\}, K_i \subseteq X^*, i = 1, 2, \dots, l, \bigcup_{i=1}^l K_i = X^*, K_i \cap K_j = \emptyset, i \neq j$. Обозначим

$D_i = \{\mathbf{x}'_i\}$ - множество всех допустимых \mathbf{x}'_i , соответствующих вектору \mathbf{x}_i (т.е.

$$x'_{ij} = \begin{cases} x_{ij}, & x_{ij} \neq \Delta, \\ \in M_j, & x_{ij} = \Delta. \end{cases}$$

Пусть $\mathbf{x}^*_t \in K_\lambda$. При замене \mathbf{x}^*_t на некоторый \mathbf{x}'_t , множество $K_\lambda \setminus \{\mathbf{x}^*_t\} \cup \{\mathbf{x}'_t\}$ может оказаться не кластером с позиций алгоритма A . Обозначим через $f_t(K)$ - долю объектов \mathbf{x}'_t из D_t , для которых множество $K_\lambda \setminus \{\mathbf{x}^*_t\} \cup \{\mathbf{x}'_t\}$ будет кластером, а также все остальные кластеры останутся без изменений.

Определение 1. Степенью определенности $f(K)$ кластеризации $K = \{K_1, K_2, \dots, K_l\}$ назовем величину $f(K) = \frac{1}{m} \sum_{t=1}^m f_t(K)$.

Рассмотрим вопрос вычисления степени определенности на примере алгоритма k -межгрупповых средних.

Пусть для выборки полных описаний $X^* = \{\mathbf{x}^*_1, \mathbf{x}^*_2, \dots, \mathbf{x}^*_m\}$ получена кластеризация $K = \{K_1, K_2, \dots, K_l\}$ с помощью метода k -внутригрупповых средних [8]. Это означает, что для $\forall \mathbf{x}^*_t \in K_i$ имеет место

$$\|\mathbf{x}^*_t - \mathbf{m}^*_i\| \leq \|\mathbf{x}^*_t - \mathbf{m}^*_j\|, \quad \forall j \neq i, \quad (1)$$

где $\mathbf{m}^*_i = \frac{1}{n_i} \sum_{\mathbf{x}^*_j \in K_i} \mathbf{x}^*_j$.

Вычисление $f_t(K)$ сводится к проверке неравенств, выполненных при замене \mathbf{x}^*_t на различные допустимые $\mathbf{x}'_t = (x'_{t1}, x'_{t2}, \dots, x'_{tm})$. Во-первых, проверяется, что объект остался ближе к своему измененному центру, чем к остальным при замене на допустимое описание:

$$\|\mathbf{x}'_t - \mathbf{m}'_i\| \leq \|\mathbf{x}'_t - \mathbf{m}^*_j\|, \quad \forall j \neq i. \quad (2)$$

Во-вторых, проверяется, что после замены одного объекта на допустимое описание остальные объекты остались ближе к своим центрам:

$$\|\mathbf{x}^*_p - \mathbf{m}^*_{i_p}\| \leq \|\mathbf{x}^*_p - \mathbf{m}'_i\|, \quad \forall p \neq t. \quad (3)$$

Неравенства (2) можно переписать в следующем виде:

$$\left\| \mathbf{x}^*_t - \Delta \mathbf{x}_t - \mathbf{m}^*_i + \frac{\Delta \mathbf{x}_t}{n_i} \right\|^2 \leq \|\mathbf{x}^*_t - \Delta \mathbf{x}_t - \mathbf{m}^*_j\|^2, \quad \text{где } \Delta \mathbf{x}_t = \mathbf{x}^*_t - \mathbf{x}'_t,$$

После элементарных преобразований имеем неравенство

$$\|\mathbf{x}^*_t - \mathbf{m}^*_i\|^2 - \|\mathbf{x}^*_t - \mathbf{m}^*_j\|^2 + 2(\Delta \mathbf{x}_t, \frac{\mathbf{x}^*_t}{n_i} - \mathbf{m}^*_j + \mathbf{m}^*_i \frac{(n_i - 1)}{n_i}) - \|\Delta \mathbf{x}_t\|^2 \frac{(2n_i - 1)}{n_i^2} \leq 0. \quad (4)$$

Неравенства (3) можно преобразовать аналогично:

$$\begin{aligned} \|\mathbf{x}^*_p - \mathbf{m}^*_{i_p}\|^2 &\leq \left\| \mathbf{x}^*_p - \mathbf{m}^*_i + \frac{\Delta \mathbf{x}_t}{n_i} \right\|^2, \\ \|\mathbf{x}^*_p - \mathbf{m}^*_{i_p}\|^2 - \|\mathbf{x}^*_p - \mathbf{m}^*_i\|^2 + 2(\Delta \mathbf{x}_t, \frac{\mathbf{m}^*_i - \mathbf{x}^*_p}{n_i}) - \|\Delta \mathbf{x}_t\|^2 \frac{1}{n_i^2} &\leq 0. \end{aligned} \quad (5)$$

Вычисление $f_t(K)$ сводится к проверке выполнения неравенств вида $a + \sum_{i \in \Omega_t} y_i c_i + b \sum_{i \in \Omega_t} y_i^2 \leq 0$,

$p + \sum_{i \in \Omega_t} y_i r_i + q \sum_{i \in \Omega_t} y_i^2 \leq 0$ где $a, b, p, q, c_i, r_i, i = 1, 2, \dots, k$ – константы для заданного t , а

$y_i \in \{x_{ii}^* - x_{ii}' : x_{ii}' \in M_i\}$. При малых k здесь возможен перебор, при больших – оценка по случайной выборке.

Аналогично можно вычислить степень определенности кластеризаций, полученных другими алгоритмами A . Для этого достаточно посчитать (оценить) долю полных описаний, которые будут соответствовать критериям останова алгоритма A .

Кластеризация неполных данных на базе построения коллективных решений.

По выборке $X = \{x_1, x_2, \dots, x_m\}$ формируется N выборок полных описаний

$X^{(i)} = \{x^{(i)}_1, x^{(i)}_2, \dots, x^{(i)}_m\}, i = 1, 2, \dots, N$, где $x^{(i)}_{ij} = \begin{cases} x_{ij}, & x_{ij} \neq \Delta, \\ \in M_j, & x_{ij} = \Delta \end{cases}$ (вероятность присвоения

$x^{(i)}_{ij}$ некоторого значения из M_j равна его частоте встречаемости на обучающей выборке). Для каждой из полученных полных выборок решается задача кластеризации на l кластеров и находятся N решений $K^{(i)} = \{K^{(i)}_1, K^{(i)}_2, \dots, K^{(i)}_l\}, i = 1, 2, \dots, N$. Далее по данным решениям строится коллективная кластеризация $K = \{K_1, K_2, \dots, K_l\}$, которая и принимается как решение задачи кластеризации с прочерками.

Определение 2. Степенью определенности $\Phi(K)$ кластеризации $K = \{K_1, K_2, \dots, K_l\}$ назовем

$$\text{величину } \Phi(K) = \sum_{i=1}^N \max_{\langle t_1, t_2, \dots, t_l \rangle} \sum_{j=1}^l |K_j \cap K_{t_j}^i| / mN.$$

Определение 3. Степенью определенности $F(K)$ кластеризации $K = \{K_1, K_2, \dots, K_l\}$ назовем

$$\text{величину } F(K) = \min_{i=1, \dots, N} \max_{\langle t_1, t_2, \dots, t_l \rangle} \sum_{j=1}^l |K_j \cap K_{t_j}^i| / m.$$

Здесь $\langle t_1, t_2, \dots, t_l \rangle$ некоторая перестановка от $\langle 1, 2, \dots, l \rangle$. Величина $\max_{\langle t_1, t_2, \dots, t_l \rangle} \sum_{j=1}^l |K_j \cap K_{t_j}^i|$

характеризует близость кластеризаций K и $K^{(i)}, i = 1, 2, \dots, N$ (при равных кластеризациях она равна m). Критерий $\Phi(K)$ характеризует нормированную среднюю близость коллективной кластеризации K относительно допустимых выборок. Критерий $F(K)$ характеризует наихудший случай.

Впервые задача построения коллективных классификаций и комитетный алгоритм ее решения были предложены в работах [9-10]. В настоящем подходе к кластеризации неполных данных использовался метод [11]. Ранее коллективное решение строилось некоторым алгоритмом по множеству кластеризаций, полученных для одной и той же выборки различными методами кластеризации. В нашем случае коллективное решение будет строиться также фиксированным алгоритмом, но по множеству кластеризаций, полученных одним методом кластеризации для различных выборок $X^{(i)} = \{x^{(i)}_1, x^{(i)}_2, \dots, x^{(i)}_m\}, i = 1, 2, \dots, N$. Опишем алгоритм.

Результаты кластеризации выборки $X^{(i)} = \{\mathbf{x}^{(i)}_1, \mathbf{x}^{(i)}_2, \dots, \mathbf{x}^{(i)}_m\}, i = 1, 2, \dots, N$, некоторым методом кластеризации можно записать в виде трехмерной информационной матрицы $\|\alpha_{ij}^v\|_{m \times l \times N}$, $\alpha_{ij}^v \in \{0, 1\}$,

$$\sum_{j=1}^l \alpha_{ij}^v = 1, i = 1, \dots, m, j = 1, \dots, l, v = 1, \dots, N.$$

Ее подматрицу $\|\alpha_{ij}^v\|_{l \times N}$, $i = 1, 2, \dots, m$ можно рассматривать как новое признаковое описание объекта \mathbf{x}'_i . В качестве коллективного решения задачи кластерного анализа принимается кластеризация выборки данных m матричных описаний.

Рассмотрим реализацию описанного подхода, на примере использования метода « k – внутригрупповых средних». На каждом итерационном шаге алгоритма нам необходимо выполнять две основные операции, а именно, нахождение центра выделенного множества объектов и нахождение расстояний между объектами и центром. Далее в качестве кластеризуемой выборки будем рассматривать выборку $\tilde{X} = \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_m\}, \tilde{\mathbf{x}}_t = \|\alpha_{ij}^v\|_{l \times N}, t = 1, 2, \dots, m$.

В качестве расстояния между объектами используется расстояние Хемминга:

$$\rho(\tilde{\mathbf{x}}_t, \tilde{\mathbf{x}}_k) = \sum_{j=1}^l \sum_{v=1}^N |\alpha_{ij}^v - \alpha_{kj}^v|. \quad (6)$$

Центральной матрицей подмножества матриц $\tilde{X}' \subseteq \tilde{X}$ будем называть матрицу $\tilde{\mathbf{x}}^* = \|y_j^v\|_{l \times N}$:

$$\tilde{\mathbf{x}}^* = \arg \min_{\tilde{\mathbf{x}} \in \tilde{X}'} \sum_{\tilde{\mathbf{x}}_i \in \tilde{X}'} \rho(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}).$$

(центральная матрица может быть не единственной). Нахождение центральной матрицы подмножества \tilde{X}' , $L = |\tilde{X}'|$ (для простоты обозначений считаем, что $\tilde{X}' = \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_L\}$) представляется в виде задачи дискретной оптимизации с Nl переменными (5-7):

$$\sum_{i=1}^L \sum_{v=1}^N \sum_{j=1}^l |y_j^v - \alpha_{ij}^v| \longrightarrow \min, \quad (7)$$

$$\sum_{j=1}^l y_j^v = 1, v = 1, \dots, N, \quad (8)$$

$$y_j^v \in \{0, 1\}, j = 1, \dots, l, v = 1, \dots, N. \quad (9)$$

Задача (7-9) распадается на N независимых оптимизационных задач

$$\sum_{i=1}^L \sum_{j=1}^l |y_j^v - \alpha_{ij}^v| \longrightarrow \min, \sum_{j=1}^l y_j^v = 1, y_j^v \in \{0, 1\}, v = 1, 2, \dots, N.$$

Обозначим $H(\mu) = \sum_{i=1}^L \left(\sum_{\substack{j=1 \\ j \neq \mu}}^l \alpha_{ij}^v + 1 - \alpha_{i\mu}^v \right)$ и пусть $\eta = \arg \min_{\mu} \Phi(\mu)$.

Тогда бинарный вектор $\mathbf{y}^{*v}, y_i^{*v} = \begin{cases} 0, & i \neq \eta \\ 1, & i = \eta \end{cases}$ и будет решением данной задачи.

Таким образом, используя метрику (6) для бинарных объектов размерности $l \times N$ и выше определенный алгоритм вычисления бинарных центральных бинарных матриц подмножеств матриц, в качестве алгоритма коллективной кластеризации может быть использован стандартный алгоритм

k – внутригрупповых средних (данный алгоритм назван в [11] алгоритмом коллективных k – внутригрупповых средних).

Результаты экспериментов на модельных и практических данных

Предложенные алгоритмы кластеризации неполных данных были апробированы на модельных примерах и двух практических задачах. В качестве модельных задач использовались выборки смеси нормальных распределений с независимыми признаками. Математические ожидания и дисперсии классов выбирались такими, чтобы результат их кластеризации совпадал с их априорной классификацией. Визуализация одного модельного примера для четырех классов (проекция многомерных данных на плоскость обобщенных признаков, см. [8,12]) приведена на рис. 1. Рассматриваемые обучающие выборки полных описаний преобразовывались в выборки с прочерками при различных уровнях неполноты данных: задается процент w неизвестных значений признаков и в каждой строке таблицы обучения по равномерному закону распределения $w\%$ признаков считались неизвестными. Далее, полученные выборки частичных описаний восстанавливались в выборки полных описаний с помощью локального алгоритма и решалась задача их кластеризации на 4 кластера. Отдельно решалась задача кластеризации неполных выборок с помощью коллективных кластеризаций. Были построены зависимости критериев

$f(K)$, $\Phi(K)$, $F(K)$, а также зависимости показателей $\varphi(K) = \max_{\langle t_1, t_2, \dots, t_l \rangle} \sum_{j=1}^l |K_j \cap K_{t_j}^*| / m$ (где

$K^* = \{K_1^*, K_2^*, \dots, K_l^*\}$ – априорная классификация исходной модельной выборки, а K – коллективная

кластеризация данных с прочерками) и $\varphi_{\text{средн.}}(K) = \max_{\langle t_1, t_2, \dots, t_l \rangle} \sum_{j=1}^l |K_j \cap K_{t_j}^*| / m$ (где K – кластеризация

выборки при замене прочерков на средние по выборке значения) от параметра w . Критерий $f(K)$

применен к двум видам кластеризаций: K^1 – кластеризации, полученные после заполнения прочерков локальным алгоритмом; K^2 – кластеризации при замене прочерков на средние по выборке значения. Параметры локального алгоритма: $\Theta = 0.4$, $N = 50$, $\varepsilon = 0.1$, $k = 5$.

На рисунках 1-3, и 4-6, показаны, соответственно, визуализации и графики модельной и двух практических задач. Задача «breast» [13] представляет выборку из 344 описаний пациентов, имеющих доброкачественное (класс №1) или злокачественное (класс №2) новообразование, для описания пациентов использовалось девять k – значных признаков. Задача «ion» [14] (распознавание сигналов в атмосфере) имела следующие характеристики: 34 числовых признака, два класса, 351 объект. Обе задачи имеют кластерную структуру, удовлетворительно совпадающую с априорной классификацией. Вид полученных зависимостей соответствует априорным ожиданиям. Критерий $F(K)$ соответствует наилучшему возможному результату и его значение быстро падает с ростом w . Графики критериев $f(K^1)$ и $f(K^2)$ близки друг к другу, что говорит о том, что метод восстановления признаков менее важен, чем выбор вида самого критерия. Тем не менее, по результатам экспериментов видно, что значения критерия $f(K^1)$ обычно выше $f(K^2)$. Это косвенно показывает, что кластеризации, полученные после заполнения прочерков локальным алгоритмом, «лучше» кластеризаций данных при замене прочерков на средние по выборке значения. На всех задачах оказываются близки показатели $\varphi(K)$ и $\varphi_{\text{средн.}}(K)$, что, по-видимому, свидетельствует о надежности «усреднений» по множеству допустимых кластеризаций, т.е. использовании коллективных решений. Как видно из графиков, значение показателя $\varphi(K)$ в целом выше значения $\varphi_{\text{средн.}}(K)$. Это показывает, что коллективное решение позволяет получить более точную кластеризацию, чем полученную при замене прочерков на средние по

выборке значения. Критерий $\Phi(K)$ представляется наиболее объективным и обоснованным, что выражается и в стабильности его зависимости от w на всех рассмотренных задачах.



Рис. 1. Смесь нормальных распределений



Рис. 2. Задача «breast»



Рис. 3. Задача «lon»

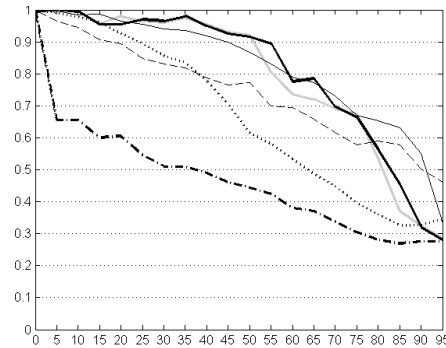


Рис. 4. Зависимости критериев и показателей от неполноты данных в модельной задаче

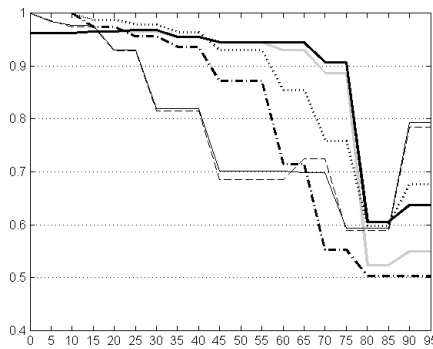


Рис. 5. Зависимости критериев и показателей от неполноты данных в задаче «breast»

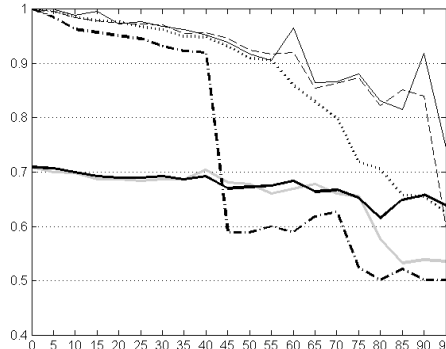


Рис. 6. Зависимости критериев и показателей от неполноты данных в задаче «lon»

Обозначения графиков критериев и показателей.

- (thin grey line) - показатель $\varphi_{\text{средн.}}(K)$
- (thick black line) - показатель $\varphi(K)$
- (solid black line) - критерий $f(K^1)$
- - - (dashed black line) - критерий $f(K^2)$
- · - · - (dash-dot black line) - критерий $F(K)$
- · · · · (dotted black line) - критерий $\Phi(K)$

Заключение

Предложенные в настоящей работе критерии степени определенности кластеризаций неполных данных основаны на оценке устойчивости полученных кластеризаций относительно возможных вариаций неполных признаков описаний. Рассмотрены различные подходы. По результатам настоящих предварительных исследований критерий $\Phi(K)$ представляется наиболее объективным. На основе предложенных критериев очевидным образом могут вычисляться и оценки степени определенности кластеризации отдельных объектов. Ранее, в работе [15] был предложен общий подход к оценке качества кластеризаций, основанный на оценке устойчивости кластеризаций относительно вариаций обучающей выборки. Представляет интерес исследование взаимосвязи рассмотренных в настоящей работе критериев степени определенности кластеризаций и критериев качества кластеризаций, а также создание критериев качества кластеризаций неполных данных. Данные вопросы будут рассмотрены в дальнейших исследованиях авторов.

В настоящее время не существует общих, универсальных алгоритмов кластеризации, тем более в случае неполноты данных. Создание новых подходов и алгоритмов расширяет возможности анализа данных, нахождения скрытых закономерностей и свойств по выборкам прецедентов.

Acknowledgements

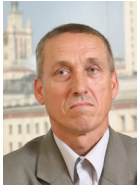
Настоящая работа выполнена при поддержке Программ Президиума РАН №14 и «Фундаментальные науки - медицине», Программы №2 Отделения математических наук РАН, проектов РФФИ 09-01-00409, 10-01-90015 Бел_а, 10-01-90419 Укр_а.

Библиография

- [1] R.J.A. Little, D.B. Rubin, *Statistical Analysis with Missing Data*. Wiley, New York, 1987.
- [2] Zloba, E. Statistical methods of reproducing of missing data / E.Zloba, I.Yatskiv // *Computer Modelling & New Technologies*. — 2002. — Vol.6, No.1 — P. 51-61.
- [3] F. Honghai, C. Guoshun, Y. Cheng, Y. Bingru and C. Yumei, *A SVM Regression Based Approach to Filling in Missing Values*, LNCS - Knowledge-Based Intelligent Information and Engineering Systems, Springer Berlin - Heidelberg, vol. 3683, 2005, 581-587.
- [4] Manish Sarkar and Tze-Yun Leong, *Fuzzy K-means Clustering with Missing Values*. Proc AMIA Symp. 2001: 588–592.
- [5] Katsuhiko Honda, and Hidetomo Ichihashi, *Linear Fuzzy Clustering Techniques With Missing Values and Their Application to Local Principal Component Analysis*, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 12, NO. 2, APRIL 2004, pp. 183-193.
- [6] Kiri Wagstaff, *Clustering with Missing Values: No Imputation Required*. In "Classification, Clustering, and Data Mining Applications". Studies in Classification, Data Analysis, and Knowledge Organization, 2004, Volume 0, Part VII, 649-658.
- [7] Михайлова Е.И., Рязанов В.В., Штаюра В.А. //Распознавание по прецедентам при наличии пропусков значений признаков. Математические методы распознавания образов: 14-я Всероссийская конференция. Владимирская обл., г.Суздаль, 21-26 сентября 2009 г.: Сборник докладов.-М.: МАКС Пресс, 2009. Стр. 163-164.
- [8] Дуда Р., Харт П., *Распознавание образов и анализ сцен*. Издательство "Мир", Москва, 1976, 511 с.
- [9] Рязанов В.В. Комитетный синтез алгоритмов распознавания и классификации // ЖВМ и МФ. 1981. Том 21, №6. С.1533-1543.
- [10] Рязанов В.В. О синтезе классифицирующих алгоритмов на конечных множествах алгоритмов классификации (таксономии) // ЖВМ и МФ, 1982. Том 22, №2. С.429-440.
- [11] Бирюков А.С., Рязанов В.В., Шмаков А.С. Решение задач кластерного анализа коллективами алгоритмов. Журнал вычислительной математики и математической физики, М.: Наука. Т.48, 2008, N 1, стр. 176-192
- [12] Ю.И.Журавлев, В.В.Рязанов, О.В.Сенько, РАСПОЗНАВАНИЕ. Математические методы. Программная система. Практические применения. Изд.во «ФАЗИС», Москва, 2006, 178 стр.

- [13] O. L. Mangasarian and W. H. Wolberg: "Cancer diagnosis via linear programming", SIAM News, Volume 23, Number 5, September 1990, pp 1 & 18.
- [14] Sigillito, V. G., Wing, S. P., Hutton, L. V., & Baker, K. B. (1989). Classification of radar returns from the ionosphere using neural networks. Johns Hopkins APL Technical Digest, 10, 262-266
- [15] Арсеев А.С., Коточигов К.Л., Рязанов В.В. : Универсальные критерии кластеризации и вопросы устойчивости. Труды 13-й Всероссийской конференции «Математические методы распознавания образов», Санкт-Петербург, 2007, с. 63-64.

Информация об авторах



Vladimir Ryazanov – Head of Department; Institution of Russian Academy of Sciences Dorodnicyn Computing Centre of RAS, Russia, 119991 Moscow, Vavilov's street, 40; e-mail: rvv@ccas.ru

Major Fields of Scientific Research: Pattern recognition, Data mining, Artificial Intelligence



Kirill Tishin – The last year student; Lomonosov Moscow State University; Faculty of Computational Mathematics and Cybernetics; Department of Mathematical Forecasting Methods, Russia, 119991, Moscow, GSP-1, Leninskie Gory, 1, p. 52; e-mail: kirill.tishin@gmail.com

Major Fields of Scientific Research: Pattern recognition, Data mining, Artificial Intelligence



Anton Schichko – The last year student; Lomonosov Moscow State University; Faculty of Computational Mathematics and Cybernetics; Department of Mathematical Forecasting Methods, Russia, 119991, Moscow, GSP-1, Leninskie Gory, 1, p. 52; e-mail: anton.schichko@gmail.com

Major Fields of Scientific Research: Pattern recognition, Data mining, Artificial Intelligence

ГРУППОВОЙ МНОГОКРИТЕРИАЛЬНЫЙ ВЫБОР ВЫЧИСЛИТЕЛЬНЫХ КЛАСТЕРОВ

Алексей Борисович Петровский, Григорий Владимирович Ройзензон

Аннотация: Рассматривается задача сравнения и выбора вычислительных кластеров, обладающих требуемыми для прикладных применений параметрами, как задача многокритериального стратегического выбора. Выбор варианта кластера производится на основе многих критериев, среди которых могут быть как количественные, так и качественные. Предложены способы формирования набора составных критериев, составляющих кластер с использованием интерактивного метода снижения размерности признакового пространства, исходя из технических и стоимостных характеристик аппаратных средств. По сконструированному таким образом составным критериям осуществляется многокритериальный выбор вычислительных кластеров с применением методов группового вербального анализа решений.

Ключевые слова: групповой вербальный анализ решений, многокритериальная порядковая классификация, агрегирование признаков, интегральные показатели оценки

ACM Classification Keywords: A.0 General Literature - Conference proceedings

Введение

В настоящее время существует достаточно много прикладных и научных задач, требующих для своего решения больших вычислительных мощностей, предоставляемых суперкомпьютерными технологиями. Высокопроизводительные вычисления необходимы в самых различных областях, в таких как: нанотехнологии, обработка потоков информации в распределенных базах данных, автоматизация проектирования, компьютерное управление производственными процессами, анализ фондового рынка, управление сотовой связью, моделирование погоды, биоинформатика, биохимия, биофизика, теплофизика, динамика жидкостей и газов, электромагнетизм, исследование генома человека, исследование прочности материалов [Еремеев, 1993], [Смирнов, Юсупов, 1997], [Янилкин и др., 2010], [Буравлёв и др., 2009] и в других.

Необходимость в разработке более дешевой аппаратной части суперкомпьютеров побудила обратить более пристальное внимание на кластерные технологии. По определению компании Digital Equipment Corporation (DEC), кластер — это группа вычислительных машин, которые связаны между собой и функционируют как один узел обработки информации [Савяк, 2002]. Активное развитие кластерных технологий обусловлено тем, что, используя стандартные компоненты — обычные процессоры, материнские платы, сетевые компоненты и тому подобное, которыми буквально «завален» весь мировой рынок, — стало возможным создание значительно более дешевых вычислительных комплексов не только не уступающих, но и порой превосходящих по производительности продукцию известных фирм, например, таких как Cray.

Кластеры можно разделить на две большие категории. Первая категория — кластеры высокой готовности или отказоустойчивые кластеры [Климанов и др., 2002]. Для таких кластерных систем на первое место выходит понятие надежности. Вторая категория — высокопроизводительные или вычислительные кластеры. Для этой категории главным является производительность. Кластеры можно также разделить на однородные (состоящие из вычислительных узлов одной и той же конфигурации) и неоднородные (состоящие из вычислительных узлов различной конфигурации). В данной работе акцент сделан на однородные вычислительные кластеры (далее ВК).

Модульная структура ВК позволяет гибко и последовательно наращивать его производительность за счет добавления новых вычислительных узлов. При этом даже морально устаревшие модули могут использоваться и в дальнейшем. Принимая во внимание многообразие процессоров, сетевых технологий,

программных продуктов для выполнения расчетов и тому подобное, перед конечным пользователем (или группой пользователей), заинтересованном (-ых) в использовании ВК для решения своих конкретных вычислительных задач, стоит достаточно непростая задача выбора конфигурации ВК.

Построение системы критериев

Проблема выбора ВК, предназначенных для решения прикладных задач пользователя, рассматривается как слабоструктурированная задача многокритериального стратегического выбора. Предварительно составляется перечень возможных вариантов, из которого будет выбран наиболее предпочтительный ВК.

Выбор ВК зависит от следующих основных факторов.

Используемая прикладная программа предъявляет определенные требования к характеристикам аппаратной платформы, что накладывает ограничения на выбор конфигурации ВК или отдельных компонентов ВК. Прикладная программа обуславливает также выбор операционной системы и кластерного программного обеспечения. Важность этого фактора связана с тем, что для решения задачи пригодны не все, а только вполне определенные прикладные программы, стоимость которых может быть на порядок больше, чем стоимость ВК.

Размерность счетной задачи и необходимый временной выигрыш позволяют определить требуемые производительность, минимальный суммарный объем и тип оперативной памяти, суммарный объем и тип дискового пространства. Это, в свою очередь, позволяет выбрать ту или иную базовую аппаратную платформу, а также влияет на выбор типа процессора.

Точность расчетов в ряде случаев также оказывает влияние на выбор типа процессора. Чем выше требуемая точность, тем больше времени нужно для расчета.

Интенсивность обмена данными между вычислительными узлами позволяет выбрать технологию построения сети ВК.

Вошедшие в перечень варианты ВК должны удовлетворять следующим основным требованиям:

- совместимость компонентов ВК с имеющимся программным обеспечением для выполнения расчетов;
- приемлемая производительность ВК;
- приемлемая стоимость ВК.

Особенность рассматриваемых объектов выбора (сложные технические системы, в частности, ВК) состоит в том, что они характеризуются большим числом показателей. Поскольку вариантов немного, то обычно все варианты несравнимы друг с другом по своим характеристикам. В этой ситуации известные методы решения задачи выбора лучшего объекта оказываются неэффективными. Поэтому прежде, чем применять какой-либо метод вербального анализа решений, необходимо решить еще одну вспомогательную, но крайне важную задачу: построить процедуру, которая позволяет агрегировать большое число базовых (технических, эксплуатационных, стоимостных) характеристик ВК в небольшое число критериев, имеющих порядковые шкалы оценок (количественные и качественные). Сокращенное описание объектов позволит упростить процедуру решения исходной задачи выбора. Выбор критериев и формирование шкал оценок проводится ЛПР самостоятельно или с привлечением системного аналитика. Предлагается следующий подход формирования набора критериев [Ройзензон, 2004], [Ройзензон, 2005].

Первоначально составляется перечень всех базовых показателей, характеризующих отдельные компоненты кластера, кластер в целом и условия его эксплуатации. Характеристики, описывающие ВК, можно представить в виде иерархической системы, нижним уровнем которой служат выделенные базовые показатели. Например, процессор характеризуется такими базовыми показателями, как архитектура (количество ядер / число поддерживаемых потоков), тактовая частота, объем кэша второго и третьего уровня, поддерживаемая частота системной шины. Некоторые из базовых показателей удобно объединять в составные показатели, которые выступают как оценки следующего уровня иерархии. После классификации эти общие оценки наполняются конкретным содержанием.

Следующим этапом является формирование вспомогательных шкал оценок для каждого базового показателя. Шкалы могут иметь числовые точечные, интервальные или вербальные (словесные) оценки. Шкалы оценок могут совпадать с обычно используемыми на практике, либо конструироваться специально для данного критерия. Например, производительность ВК оценивается в Гфлопсах, стоимость - в млн.

рублей. Для сокращения размерности описания объекта часто бывает удобно перейти от непрерывной шкалы оценки к дискретной шкале, имеющей небольшое число оценок на шкале, и от количественной шкалы к качественной шкале. Например, можно оценивать стоимость оценками «низкая», «средняя», «высокая», указав для каждой из оценок соответствующие интервалы величин. Все сформированные оценки ЛПР упорядочивает от лучшей к худшей.

Далее ЛПР по своему усмотрению определяет число и состав критериев, их содержание. В качестве критерия можно выбрать один из базовых показателей (например, производительность, стоимость) или несколько характеристик, объединенных в составной критерий. ЛПР устанавливает, какие технические и эксплуатационные показатели будут считаться самостоятельными критериями, а какие будут отнесены к тому или иному составному критерию. Шкалы простых критериев, являющихся базовыми показателями, уже построены на предыдущем этапе. Для формирования шкал оценок по составным критериям можно воспользоваться несколькими процедурами. При формировании шкалы оценок составного критерия важно учесть, что одна часть характеристик, входящих в состав подобного критерия, может рассматриваться как самостоятельная, а другая часть характеристик может быть составной. Поэтому процедура построения шкалы составного критерия сама может состоять из нескольких этапов.

На основе анализа литературы [Авен, 1981], [Иванилов, 2003], [Климанов и др., 2002], [Ройзензон, 2005], [Рытиков и др., 1988], [Фортов и др., 2002], [Буравлёв и др., 2009], [Янилкин и др., 2010] был предложен список характеристик ВК, которые разделены на несколько групп. К первой отнесем технические характеристики кластера (процессор, базовая аппаратная платформа, технология построения сети, оперативная память, дисковая память).

Используемая операционная система и кластерное программное обеспечение составляют вторую группу. Производными от них является третья группа показателей: производитель технических и программных средств; производительность кластера; стоимость кластера (группа «Обобщенные характеристики»). И, наконец, в последнюю группу «Эксплуатационные характеристики» входят: энергопотребление, тепловыделение, уровень шума и условия окружающей среды.

Построение составных критериев

Исходя из базовых характеристик ВК, были предложены четыре критерия оценки ВК и разработаны шкалы их оценок. В качестве самостоятельных критериев для сравнения ВК были выбраны две базовые характеристики: стоимость и производительность кластера. Двумя другими являются составные критерии: возможность модернизации и сложность эксплуатации кластера, сконструированные из базовых характеристик и составных критериев нижнего уровня. Для построения шкал составных критериев использовался интерактивный метод снижения размерности признакового пространства [Петровский, Ройзензон, 2008б]. Рассмотрим особенности шкал составных критериев.

Возможность модернизации кластера определяется техническими и функциональными характеристиками процессора, базовой аппаратной платформы, технологией построения сети, оперативной памяти и дисковой памяти, которые являются базовыми характеристиками ВК. Возможность модернизации также напрямую связана с понятием ремонтпригодности.

Порядковая шкала оценок по критерию «Возможность модернизации кластера», сконструированная методом стратификации кортежей [Петровский, Ройзензон, 2008а], включает в себя следующие оценки:

Высокая. Вычислительные узлы оснащены перспективным процессором, перспективными базовыми аппаратными платформами. ВК использует перспективную технологию построения сети. Общий объем и тип оперативной памяти отвечает требованиям задачи и имеется возможность ее увеличения в будущем. Общий объем и тип дисковой памяти отвечает требованиям задачи и имеется возможность ее увеличения в будущем.

Средняя. Вычислительные узлы оснащены современными процессорами, современными базовыми аппаратными платформами. Технология построения сети отвечает текущим требованиям. Общий объем и тип оперативной памяти отвечает требованиям расчетной задачи, однако большого резерва по наращиванию ее объема нет. Общий объем и тип дисковой памяти отвечает требованиям расчетной задачи, однако большого резерва по наращиванию ее объема нет.

Низкая. Вычислительные узлы оснащены устаревающими процессорами, устаревающими базовыми аппаратными платформами. ВК использует устаревающую технологию построения сети. Нарастивание объема оперативной памяти ограничено. Нарастивание объема дисковой памяти ограничено.

Сложность эксплуатации ВК во многом определяется характеристиками производителей технических и программных средств, а также дополнительными затратами, необходимыми для организации условий эксплуатации ВК.

Таким образом, в составной критерий «Сложность эксплуатации кластера» входят базовые характеристики: производитель, стоимость владения и составной критерий «Эксплуатационные характеристики».

Шкала составного критерия «Эксплуатационные характеристики» была сформирована с использованием метода порядковой классификации ОРКЛАСС [Ларичев, 2006]. В составной критерий «Эксплуатационные характеристики» вошли следующие базовые показатели: энергопотребление, тепловыделение, уровень шума, требования к условиям окружающей среды. Для каждого из представленных базовых показателей ЛПР определил соответствующие порядковые вербальные шкалы оценок, которые были получены путем преобразования исходных интервальных шкал. Например:

K_1 – энергопотребление (0 – низкое, 1 – среднее, 2 – высокое); K_2 – тепловыделение (0 – низкое, 1 – среднее, 2 – высокое); K_3 – уровень шума (0 – низкий, 1 – средний, 2 – высокий); K_4 – условия окружающей среды (0 – повышенные требования к условиям окружающей среды не предъявляются, 1 – кластер предъявляет повышенные требования к условиям окружающей среды).

Для составного критерия «Эксплуатационные характеристики» была предложена вспомогательная шкала, состоящая из 3-х порядковых вербальных оценок: (D_1 – вычислительный кластер не предъявляет серьезных требований по организации условий эксплуатации, D_2 – вычислительный кластер предъявляет незначительные требования по организации условий эксплуатации, D_3 – вычислительный кластер предъявляет повышенные требования по организации условий эксплуатации).

В результате была построена следующая полная непротиворечивая порядковая классификация эксплуатационных характеристик ВК, представленная ниже.

Класс D_1 (верхняя граница: 0000; нижняя граница: 0000); Класс D_2 (верхняя граница: 1000,0100,0010,0001; нижняя граница: 2210,2120,2001,1101,0201,0111,0021); Класс D_3 (верхняя граница: 0220,2101,1201,1011,0211,0121; нижняя граница: 2221).

Порядковая шкала оценок по составному критерию «Сложность эксплуатации», сконструированная методом стратификации кортежей, включает в себя следующие оценки:

Низкая. Вычислительная техника представлена всемирно известным производителем. ВК не предъявляет серьезных требований по организации условий эксплуатации. ВК обладает низкой стоимостью владения.

Средняя. Вычислительная техника представлена достаточно хорошо известным производителем. ВК предъявляет незначительные требования по организации условий эксплуатации. ВК обладает средней стоимостью владения.

Высокая. Вычислительная техника представлена малоизвестным производителем. ВК предъявляет повышенные требования по организации условий эксплуатации. ВК обладает высокой стоимостью владения.

Пример

Рассмотрим иллюстративный пример решения задачи многокритериального выбора наилучшего ВК, в котором используются сконструированные составные критерии. После перехода от числовых или вербальных оценок базовых показателей к критериальным оценкам может случиться так, что варианты ВК станут сравнимыми и, более того, некоторый вариант (или некоторые) окажется наилучшим. Если же наилучший вариант сразу выделить нельзя, то для его нахождения можно воспользоваться одним из методов вербального анализа решений. В этом случае размерность описания такой новой задачи многокритериального выбора и сложность ее решения будут существенно меньше исходной.

Предположим, что перед ЛПР стоит задача выбора ВК из трех конфигураций, представленных различными производителями вычислительной техники. Оценки ВК по предложенному набору составных критериев приведены в таблице 1.

Таблица 1. Оценки вычислительных кластеров по критериям

№	Вычислительный кластер / Критерий	№ 1	№ 2	№3
А.	Стоимость, млн. рублей	0,6	0,45	0,3
Б.	Производительность, Гфлопс	150	120	90
В.	Возможность модернизации	Высокая	Средняя	Низкая
Г.	Сложность эксплуатации	Средняя	Средняя	Высокая

Как следует из таблицы 1, все варианты ВК теперь сравниваются по гораздо меньшему набору показателей (по четырем критериям), однако по-прежнему остаются несравнимыми. Для выбора лучшего ВК можно воспользоваться методом вербального анализа решений ЗАПРОС [Ларичев, Мошкович, 1996], [Ларичев, 2006], установив субъективную предпочтительность оценок по разным критериям и построив единую порядковую шкалу. Возможность сформировать разные наборы критериев позволяет сравнить полученные результаты выбора для разных вариантов с целью оценки качества выбора. В ряде случаев для надежности выбора целесообразно проранжировать варианты ВК сразу несколькими методами (например, ПАРК [Ларичев, Мошкович, 1996], АНР [Саати, 1993]) и сопоставить результаты.

Групповой выбор

Задача выбора ВК существенно осложняется в ситуации, когда присутствуют несколько ЛПР (экспертов). В этом случае один и тот же объект (ВК) может существовать в нескольких версиях или копиях с разнообразными, несогласованными или противоречивыми значениями признаков, а свертывание числовых и/или вербальных оценок является или математически некорректным или невозможным. При наличии многих признаков и нескольких экспертов или ЛПР, необходимо агрегировать индивидуальные ранжировки по отдельным признакам, которые, как правило, не совпадают. В рассмотренном примере пять ЛПР проранжировали три варианта конфигураций ВК тремя различными методами многокритериального выбора (ЗАПРОС [Ларичев, 2006], ПАРК [Ларичев, Мошкович, 1996], АНР [Саати, 1993]) по предложенному набору составных критериев.

Для групповой сортировки многопризнаковых объектов (ВК) использован метод АРАМИС (Агрегирование и Ранжирование Альтернатив около Многопризнаковых Идеальных Ситуаций) [Петровский, 2009], который, в общем случае, не требует предварительного построения индивидуальных ранжировок и позволяет упорядочивать многопризнаковые объекты на основе противоречивых предпочтений нескольких ЛПР. Многопризнаковые объекты были представлены как мультимножества [Петровский, 2003] и упорядочивались по показателю близости к «идеальной» точке в метрическом пространстве мультимножеств.

В методе АРАМИС объекты описываются множеством повторяющихся количественных и/или качественных атрибутов Q_1, \dots, Q_m представленных в виде мультимножеств. Многопризнаковые объекты A_1, \dots, A_n рассматриваются как точки метрического пространства мультимножеств (A, d) с различными метриками d [Петровский, 2003], которые можно сравнивать и упорядочивать по показателю относительной близости к наилучшему (идеальному) объекту A_{\max} или наихудшему (антиидеальному) A_{\min} в этом пространстве. Наилучший и наихудший объекты (которые могут быть гипотетическими) имеют наилучшие и наихудшие оценки по всем критериям Q_s . Объекты упорядочиваются по показателю относительной близости к наилучшему объекту $I^{\max}(A_i) = d(A_{\max}, A_i) / [d(A_{\max}, A_i) + d(A_{\min}, A_i)]$, где $d(A_{\max}, A_i)$ расстояние до наилучшего объекта A_{\max} и $d(A_{\min}, A_i)$ расстояние до наихудшего объекта A_{\min} .

Рассмотрим методы M_1 (ЗАПРОС), M_2 (ПАРК), M_3 (АНР), использованные ЛПР для ранжирования ВК, как новые атрибуты, которые характеризуют ВК. Каждый атрибут M_j имеет оценки m_j^1, m_j^2, m_j^3 , соответствующие порядковым номерам ВК в ранжировках, полученным ЛПР (т.е. m_j^1 означает, что методом M_j объект имеет в ранжировке порядковый номер 1 и т.д.). Теперь каждый ВК A_i можно

представить как множество с повторяющимися элементами или мультимножество A_i , порожденное множеством методов $X=M_1 \cup M_2 \cup M_3$:

$$A_i = \{k_{A_i}(m_1^1) \circ m_1^1, \dots, k_{A_i}(m_1^3) \circ m_1^3; \dots; k_{A_i}(m_3^1) \circ m_3^1, \dots, k_{A_i}(m_3^3) \circ m_3^3\}.$$

Здесь $k_{A_i}(m_j^{h_j})$, $h_j=1, \dots, 3$, $j=1, \dots, 3$ показатель кратности, соответствующий тому, сколько значений атрибута встречается в мультимножестве A_i , который показывает, сколько порядковый номер ВК в ранжировке $m_j^{h_j}$ получен всеми ЛПР во время ранжирования соответствующим методом M_j ; знак \circ обозначает, сколько $k_{A_i}(m_j^{h_j})$ раз атрибут $m_j^{h_j}$ встречается в описании объекта A_i .

Соответственно, ВК A_1 , A_2 и A_3 представляются в виде следующих мультимножеств:

$$A_1 = \{3 \circ m_1^1, 2 \circ m_1^2, 0 \circ m_1^3; 2 \circ m_2^1, 2 \circ m_2^2, 1 \circ m_2^3; 2 \circ m_3^1, 2 \circ m_3^2, 1 \circ m_3^3\},$$

$$A_2 = \{2 \circ m_1^1, 2 \circ m_1^2, 1 \circ m_1^3; 2 \circ m_2^1, 1 \circ m_2^2, 2 \circ m_2^3; 1 \circ m_3^1, 3 \circ m_3^2, 1 \circ m_3^3\},$$

$$A_3 = \{0 \circ m_1^1, 1 \circ m_1^2, 4 \circ m_1^3; 1 \circ m_2^1, 2 \circ m_2^2, 2 \circ m_2^3; 2 \circ m_3^1, 0 \circ m_3^2, 3 \circ m_3^3\},$$

наилучший ВК A_{\max} и наихудший ВК A_{\min} представлены мультимножествами:

$$A_{\max} = \{5 \circ m_1^1, 0, 0; 5 \circ m_2^1, 0, 0; 5 \circ m_3^1, 0, 0\}, A_{\min} = \{0, 0, 5 \circ m_1^3; 0, 0, 5 \circ m_2^3; 0, 0, 5 \circ m_3^3\}.$$

Расстояния между отдельными мультимножествами (ВК) A_i , наилучшим (идеальным) A_{\max} и наихудшим (антиидеальным) мультимножествами A_{\min} , полученные после обработки результатов, равны соответственно: $d_1^{\max}(A_1)=16$, $d_1^{\max}(A_2)=20$, $d_1^{\max}(A_3)=24$; $d_1^{\min}(A_1)=26$, $d_1^{\min}(A_2)=22$, $d_1^{\min}(A_3)=12$, где d_1 — метрика типа Хемминга. Значения показателя $I^*(A_i)$ относительной близости ВК A_i к наилучшему ВК A_{\max} задаются величинами: $I^{\max}(A_1)=0,381$, $I^{\max}(A_2)=0,4762$, $I^{\max}(A_3)=0,6667$. Итоговое упорядочение ВК по увеличению показателя относительной близости к наилучшему ВК A_{\max} имеет вид: $A_1 > A_2 > A_3$.

Заключение

В работе предложен «прозрачный» подход индивидуального и группового выбора ВК. Выбор ВК рассматривается как слабоструктурированная задача многокритериального стратегического выбора. Составные критерии при этом формируются индивидуально или несколькими ЛПР (экспертами) по многим числовым, символьным или вербальным критериям. Используя интерактивную процедуру снижения размерности признакового пространства, сконструированы составные критерии «Возможность модернизации» и «Сложность эксплуатации» ВК путем агрегирования более простых критериев (базовых характеристик), например технических, эксплуатационных, стоимостных и т.п. Важной особенностью процедуры является возможность сформировать разные наборы критериев, с тем, чтобы сравнить полученные результаты для разных вариантов с целью оценки качества выбора. Для индивидуального выбора ВК по составным критериям использован метод вербального анализа решений ЗАПРОС. Используя метод АРАМИС для группового упорядочивания многопризнаковых объектов, основанный на теории метрических пространств мультимножеств, были построены ранжировки ВК с использованием комбинации различных методов многокритериального выбора (ЗАПРОС, ПАРК, АНР). Подход теории мультимножеств позволяет обнаруживать, представлять и использовать доступную информацию, анализировать полученные результаты и их особенности, особенно для несогласованных многокритериальных оценок ВК и противоречивых предпочтений ЛПР.

Благодарности

Работа поддержана программами фундаментальных исследований президиума РАН «Интеллектуальные информационные технологии, математическое моделирование, системный анализ и автоматизация» и ОНИТ РАН «Информационные технологии и методы анализа сложных систем», Российским фондом фундаментальных исследований (проекты 09-07-00009, 11-07-00230, 11-07-00398).

Список литературы

[Авен, 1981] Авен О. И. Что же такое АСУ?— М.: Наука, 1981.— 175 с.

[Буралёв и др., 2009] Масштабируемые мультипроцессорные вычислительные системы высокой производительности / Буралёв А., Чельдиев М., Барыбин А., Костенко В., Тумакин Д., Петрова Г. // Современные технологии автоматизации. — 2009. — № 3. — С. 72-76.

- [Еремеев, 1993] Еремеев А. П. Организация параллельных вычислений на основе моделей потока данных // Известия РАН. Техническая кибернетика. — 1993.— № 3.— С. 212–225.
- [Иванилов, 2003] Иванилов Е. И. Некоторые аспекты выбора серверов // Корпоративные системы. — 2003. — № 4.— С. 34–36.
- [Климанов и др., 2002] Климанов В. П., Сутягин М. В., Быстрикова В. А. Задачи катастрофоустойчивости кластерных вычислительных систем // Новости искусственного интеллекта.— 2002.— Т. 3, № 51.— С. 22–30.
- [Ларичев, 2006] Ларичев О.И. Вербальный анализ решений / Под ред. А.Б.Петровского. — М.: Наука, 2006.
- [Ларичев, Мошкович, 1996] Ларичев О. И., Мошкович Е. М. Качественные методы принятия решений.— М.: Физматлит, 1996.— 208 с.
- [Петровский, 2003] Петровский А.Б. Пространства множеств и мультимножеств. – М: Едиториал УРСС, 2003.
- [Петровский, Ройзензон, 2008а] Петровский А. Б., Ройзензон Г. В. Снижение размерности признакового пространства в задачах многокритериальной классификации: стратификация кортежей // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием (КИИ-2008). Труды конференции. — Т. 2.— М.: Ленанд, 2008. — С. 262–270.
- [Петровский, Ройзензон, 2008б] Петровский А. Б., Ройзензон Г. В. Снижение размерности признакового пространства в задачах многокритериальной классификации // Decision Making and Business Intelligence Strategies and Techniques / Ed. by K. Markov, K. Ivanova, I. Mitov. — No. 3. — Sofia: FOI ITHEA, 2008.— Pp. 81–86.
- [Петровский, 2009] Петровский А. Б. Теория принятия решений. — М.: Издательский центр «Академия», 2009.
- [Ройзензон, 2004] Ройзензон Г. В. Выбор вычислительных кластеров на основе анализа количественной и качественной информации // Искусственный интеллект.— Т. 2.— Донецк, Украина: Наука і освіта, 2004. — С. 375–379.
- [Ройзензон, 2005] Ройзензон Г. В. Многокритериальный выбор вычислительных кластеров // Методы поддержки принятия решений: Сборник трудов Института системного анализа Российской академии наук / Под ред. С. В. Емельянова, А. Б. Петровского.— М.: Едиториал УРСС, 2005. — Т. 12. — С. 68–94.
- [Рытиков и др., 1988] Рытиков А. М., Ройтман Е. Я., Шафрин Ю. А. Что мешает эффективному внедрению типовых АСУ? // Цветные металлы. — 1988. — № 1. — С. 98–101.
- [Саати, 1993] Саати Т. Принятие решений. Метод анализа иерархий.— М.: Радио и связь, 1993.— 278 с.
- [Савяк, 2002] Савяк В. Эффективные кластерные решения [Электронный ресурс]. — 2002. — Режим доступа: www.ixbt.com/cpu/clustering.shtml.
- [Смирнов, Юсупов, 1997] Смирнов А. В., Юсупов Р. М. Технология параллельного проектирования: основные принципы и проблемы внедрения // Автоматизация проектирования. — 1997. — № 2. — С. 50–55.
- [Фортов и др., 2002] Создание и применение системы высокопроизводительных вычислений на базе высокоскоростных сетевых технологий / В. Е. Фортов, Г. И. Савин, В. К. Левин и др. // Информационные технологии и вычислительные системы. — 2002. — № 1. — С. 3–9.
- [Янилкин и др., 2010] Применение суперкомпьютеров для молекулярно-динамического моделирования процессов в конденсированных средах / Янилкин А.В., Жиляев П.А., Куксин А.Ю., Норман Г.Э., Писарев В.В., Стегайлов В.В. // Вычислительные методы и программирование. — 2010. — Т. 11. — С. 111-116.

Сведения об авторах

Петровский Алексей Борисович – д.т.н., заведующий лабораторией Института системного анализа РАН, Россия, Москва 117312, пр-т 60-летия Октября, 9, тел. (499)135-8503, e-mail: pab@isa.ru

Ройзензон Григорий Владимирович – к.т.н., старший научный сотрудник Института системного анализа РАН, Россия, Москва 117312, пр-т 60-летия Октября, 9, тел. (499)135-8503, e-mail: rgv@isa.ru

ПРОГНОЗИРОВАНИЕ ДИНАМИКИ ПОПОЛНЕНИЯ ФОНДОВ УГЛЕВОДОРОДОВ

Михаил Стернин, Геннадий Шепелев

Abstract: Development of rational strategies of supplementing reserves in a system of oil funds is an actual problem. A model that may assist to oil experts elaborate such strategy for system of oil fields, which are on different stages of exploration, is proposed. The model is a dynamical model of material balance with one-side cross flow from poorly studied funds to more studied ones accompanied by losses as described by transition coefficients given by experts with uncertainty. In the framework of the model is supposed that for each year of planning period levels of oil production (production program) are exogenous variables. Number of unknown variables (that are values of supplements in each fond for each year) in linear equations of the model is equal to production of quantity of time steps in planning period and number of fund types. Number of equations is much less than number of unknown variables. To solve the model it is complemented by criterion of optimality. The criterion is minimum of discounted costs needed for transitions. Coefficients in equations are interval expert estimations. It's rather hard to analyze the problem by Monte-Carlo method because this way requires large-scale calculations for each time step from planning period. One is showed that simpler equivalent model may be developed that permit to calculate needed probability distributions more easily. Besides the model has property of scaling invariance that permits to extend these distributions on further time steps.

Keywords: model of supplementing oil funds, expert analysis, property of scaling

ACM Classification Keywords: I.6.5 Model Development. I.6.6 Simulation Output Analysis. I.2.8 Plan Execution, Formation and Generation.

Введение

Несмотря на технологические достижения последних лет, современная экономика по-прежнему остается экономикой, основанной на использовании нефти и газа. В то же время многие разрабатываемые большие нефтяные месторождения находятся на стадии истощения, а новые поля не столь продуктивны, как прежние.

Процессы подготовки новых нефтяных месторождений к эксплуатации достаточно время и капиталоемкие, они требуют глубокого предварительного анализа, выбора и геолого-экономического обоснования стратегии перевода менее изученных полей в более изученные объекты, доступные для рациональной добычи.

Один из инструментов упомянутого геолого-экономического анализа – моделирование процессов пополнения фондов нефти, находящихся на различных стадиях изученности. Результаты моделирования могут помочь экспертам различного уровня прогнозировать ожидаемые объемы дефицита нефти из хорошо разведанных полей и возможные моменты его возникновения, а также оценить ожидаемые объемы работ и требуемые капиталовложения, необходимые для пополнения фондов и обеспечивающие выполнение планируемой программы добычи. Основные показатели таких планов содержатся, например, в энергетической программе России, где имеются ориентировочные цифры годовой добычи на 20-ти летнюю перспективу, а также в стратегических планах крупнейших нефтяных компаний.

Проблемная ситуация для целей моделирования выглядит следующим образом. Задана совокупность нефтяных полей. Все они имеют примерно одинаковые жизненные циклы. В сильно агрегированном виде в рамках одного нефтяного поля картина выглядит следующим образом. Имеются нефтеносные участки, находящиеся на ранних стадиях геологической изученности (они содержат ресурсы категории D , или

фонды F_D); участки, находящиеся на промежуточных стадиях изученности (с резервами и ресурсами категории C , или фонды F_C); и, наконец, хорошо изученные территории с запасами нефти промышленных категорий (фонды F_A). На самом деле структура фондов гораздо более детальная, чем здесь представленная. Однако, для целей моделирования указанного агрегирования вполне достаточно.

В процессе геологических исследований часть запасов слабо изученных фондов переходит в запасы более изученных фондов, как правило, с некоторыми потерями объемов запасов по сравнению с их оценкой в менее «зрелых» фондах. Эти потери описываются коэффициентами перевода K_{ji} . Для каждой пары непосредственно соседствующих фондов $F_i \rightarrow F_j$ задается свой коэффициент перевода. Коэффициент перевода для фондов промышленных категорий K_A - это коэффициент извлечения нефти (КИН). Он показывает, какую долю нефти, находящейся под земной поверхностью на разрабатываемых участках поля, целесообразно поднять на поверхность при имеющихся технологических, геологических и экономических условиях.

Значения коэффициентов перевода в известной мере зависят от времени, поскольку определяются знаниями о различных участках поля. Однако в реальности их считают постоянными, фиксированными на момент прогнозирования. Значения коэффициентов перевода известны, как правило, как экспертные оценки, иногда точечные, иногда интервальные. Можно видеть, что фонды образуют взаимосвязанную систему фондов.

Описанную проблемную ситуацию можно представить на следующей схеме (рис. 1).

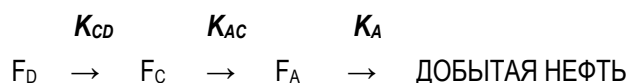


Рис. 1. Укрупненная структура системы нефтяных фондов различной изученности

Отправляясь от описанной проблемной ситуации, нами с привлечением экспертов, геологов и нефтяников-технологов, разработана следующая модель пополнения фондов нефти, обеспечивающего выполнение заданной производственной программы [Стернин и др., 2007].

Модель пополнения системы взаимосвязанных фондов нефти, находящихся на различных стадиях изученности (детерминированная постановка)

Итак, имеются фонды нефти F_A , F_C , F_D разной степени изученности, для которых известны оценки содержащихся в них запасов соответствующих категорий на момент начала прогнозирования $F_A(0)$, $F_C(0)$, $F_D(0)$. Экзогенно заданы горизонт прогнозирования T и программа добычи нефти с разбивкой по годам $Q(t)$, $t = 1, 2, \dots, T$. Обычно рассматриваемый горизонт прогнозирования составляет 15 – 20 лет. Считаются известными также желаемые в соответствии со стратегическими целями объемы фондов на конец периода прогнозирования $F_A(T)$, $F_C(T)$, $F_D(T)$, а также коэффициенты перевода K_{ji} из фонда в фонд. Тогда простейшая модель динамики пополнения фондов может быть представлена как следующая модель материального баланса (1 – 3) с перетоками между фондами:

$$F_A(t) = F_A(t-1) + F_A^+(t) - \frac{Q(t)}{K_A(t)} \quad (1)$$

$$F_C(t) = F_C(t-1) + F_C^+(t) - \frac{F_A^+(t)}{K_{AC}(t)} \quad (2)$$

$$F_D(t) = F_D(t-1) + F_D^+(t) - \frac{F_C^+(t)}{K_{CD}(t)} \quad (3)$$

Искомые величины в модели суть $F_A^+(t)$, $F_C^+(t)$, $F_D^+(t)$. В модель вводятся также ограничения на возможные объемы пополнения каждого i -го фонда ($i \in I = \{F_A, F_C, F_D\}$), такие что $F_{Li}^+(t) \leq F_i^+(t) \leq F_{Hi}^+(t)$. Эти

ограничения обусловлены производственными и финансовыми возможностями при освоении фондов. Рекомендованные экспертами значения коэффициентов перевода приведены в таблице 1, а $K_A = \text{КИН} = 0.3$.

Таблица 1. Значения коэффициентов перевода

Тип фонда	Коэффициенты перевода
F_A	K_A (КИН): 0.35
F_C	K_{AC} ($F_C \Rightarrow F_A$): 0.75
F_D	K_{CD} ($F_D \Rightarrow F_C$): 0.56

Так как число неизвестных в модели больше числа уравнений, для нахождения искомых величин может быть использован оптимизационный подход с критерием оптимальности, выбранным, например, как минимум приведенных (дисконтированных) затрат на работы по пополнению системы фондов, обеспечивающие выполнение производственной программы $Q(t)$. Важен также вопрос выбора стратегии пополнения. Она зависит от целей планирующего органа и исходного состояния фондов. Вероятно, наилучшей с геологической точки зрения, но достаточно дорогостоящей, представляется стратегия, при которой текущие объемы пополнения, пересчитанные на запасы нефти, должны быть не меньше, чем объемы добычи за тот же период. В реальной модели по решению экспертов использовалась стратегия, согласно которой текущие объемы фондов с учетом пополнения должны быть не менее чем их конечные объемы: $F_i(t) \geq F_i(T)$.

В рамках этой детерминистской модели неопределенность может быть учтена вариацией значений коэффициентов перевода. Это, однако, паллиативный подход. Для более глубокого анализа неопределенности будут использованы другие средства. Но прежде отметим, что для экспресс-анализа текущего потенциала системы фондов эксперты могут использовать агрегированные оценки, являющиеся следствием уравнений материального баланса. Модели, построенные на базе этих уравнений, назовем «экспресс-моделями».

Пусть пополнение фондов D не производится. Количество нефти, которую можно добыть из фондов промышленных категорий F_A за период T (с учетом необходимости иметь объемы $F_A(T)$ в конечном состоянии), равно $Q_A(0) = K_A[F_A(0) - F_A(T)]$, ($F_A(0) \geq F_A(T)$). Совместные возможности фондов F_A и F_C составляют $Q_{A+C}(0) = Q_A(0) + K_A K_{AC}[F_C(0) - F_C(T)]$, ($F_C(0) \geq F_C(T)$). Наконец, $Q_{A+C+D}(0) = Q_{A+C}(0) + K_A K_{AC} K_{CD}[F_D(0) - F_D(T)]$, ($F_D(0) \geq F_D(T)$). Величины $Q_A(0)$, $Q_{A+C}(0)$, $Q_{A+C+D}(0)$ показывают, сколько нефти может быть добыто из изолированно рассматриваемых подсистем « A », « $A + C$ », « $A + C + D$ » соответственно при условии перевода фондов в заданные конечные состояния $F_A(T)$, $F_C(T)$, $F_D(T)$.

Анализируя потенциал фондов промышленных категорий, без учета их пополнения из F_C , имеем $Q_A(t) = Q_A(t-1) - Q(t)$. Как только величина $Q_A(t)$ становится отрицательной, это означает, что исходных возможностей фондов промышленных категорий хватает лишь для выполнения производственной программы $Q(t)$ до года $t - 1$ включительно. При анализе совместного потенциала фондов F_A и F_C с возможностью пополнения F_A из F_C , но без пополнения из F_D , получаем: $Q_{A+C}(t) = Q_{A+C}(t-1) - Q(t)$. Как только величина $Q_{A+C}(t)$ становится отрицательной, это означает, что совместных исходных запасов фондов F_A и F_C хватает лишь для выполнения производственной программы $Q(t)$ до года $t - 1$ включительно. Аналогичные соображения применимы и для анализа исходного потенциала всей системы фондов при возможности (односторонних) перетоков из фонда в фонд.

Имеется еще одна возможность быстрой проверки текущего состояния фондов. Если обозначить накопленную (суммарную) добычу за период прогнозирования в соответствии с производственной программой через $Q(1, T)$, то при $Q_{A+C+D}(0) \geq Q(1, T)$ производственная программа может быть выполнена без пополнения системы фондов извне, однако перетоки внутри системы могут оказаться необходимыми.

При $Q_{A+C}(0) \geq Q(1, T)$ можно обойтись только ресурсами подсистемы $F_A + F_C$ с возможными перетоками внутри нее и т.д. Видно, что на этом пути проверяются лишь интегральные возможности системы фондов и различных его подсистем без расчета динамики требуемого пополнения фондов. Однако, как показано далее, на базе этой идеи может быть разработан простой алгоритм расчета динамики пополнения фондов, требуемого в соответствии с производственной программой, который обходит необходимость решения оптимизационной задачи.

Влияние неопределенности. Алгоритм расчета траекторий пополнения. Масштабная инвариантность

Обратимся теперь к исследованию той же модели, но в условиях интервально-вероятностной неопределенности коэффициентов модели. Это необходимо сделать, поскольку значения коэффициентов перевода задаются на практике экспертно или по аналогии в виде интервальных оценок, размах которых увеличивается при переходе к фондам, находящимся на начальных стадиях изученности. Рекомендуемые экспертами интервалы значений этих коэффициентов приведены в таблице 2 [Кноринг, 1992].

Таблица 2. Интервальные оценки коэффициентов перевода

Тип фонда	Коэффициенты перевода
F_A	K_A (КИН): 0.3 - 0.4
F_C	K_{AC} ($F_C \Rightarrow F_A$): 0.6 - 0.9
F_D	K_{CD} ($F_D \Rightarrow F_C$): 0.125 - 1.0

Таким образом, значения всех коэффициентов перевода лежат в интервале от минимального значения (K_m), пессимистический сценарий, до максимального значения (K_M), оптимистический сценарий. Решая оптимизационные задачи для указанных сценариев, можно получить граничные траектории для требуемого пополнения всех фондов. Внутренняя часть этих трубок содержит все возможные (в рамках ограничений) траектории пополнения для различных сочетаний значений коэффициентов перевода. Чтобы изучить эту область, надо построить распределения вероятностей для объемов пополнения каждого фонда в каждом году периода прогнозирования. Естественно использовать для этого метод статистического моделирования, но способ, состоящий в решении оптимизационной задачи для каждой истории Монте-Карло не слишком рационален. Есть ли другая возможность решения задачи оценки динамики пополнения фондов?

Эта возможность связана с использованием оптимизационной стратегии «точно в срок» («Just in Time Strategy») и экспресс-моделей предыдущего раздела. Согласно этой стратегии надо с самого начала использовать возможности фондов промышленных категорий, затем, когда накопленная добыча превзойдет их, добавить потенциал фонда F_C , затем фонда F_D и только потом приступить к пополнению фондов извне.

Рассмотрим этот подход более подробно. Минимальные, по наличию нефти, возможности фондов F_A , F_C , F_D , связанные с реализацией минимальных значений надлежащих коэффициентов перевода, равны, соответственно, $Q_{Am}(0) = K_{Am}[F_A(0) - F_A(T)]$, $Q_{Cm}(0) = K_{Am}K_{ACm}[F_C(0) - F_C(T)]$, $Q_{Dm}(0) = K_{Am}K_{ACm}K_{CDm}[F_D(0) - F_D(T)]$. При сопоставлении этих величин с требуемой суммарной добычей за период прогнозирования $Q(1, T)$ имеем четыре возможности: а) $Q_{Am}(0) \geq Q(1, T)$, б) $Q_{Am}(0) + Q_{Cm}(0) \geq Q(1, T) \geq Q_{Am}(0)$, в) $Q_{Am}(0) + Q_{Cm}(0) + Q_{Dm}(0) \geq Q(1, T) \geq Q_{Am}(0) + Q_{Cm}(0)$, г) $Q(1, T) \geq Q_{Am}(0) + Q_{Cm}(0) + Q_{Dm}(0)$.

Рассмотрим ситуацию г). Пусть на интервале $[K_{Am}, K_{AM}]$ экспертом задано распределение вероятностей. Разыграем значение K_A и найдем величину $Q_A(0)$, отвечающую этому значению коэффициента перевода. Если накопленная к окончанию года t добыча $Q(1, t)$ не превосходит $Q_A(0)$, то план добычи,

предусмотренной производственной программой для всех лет, включая t , может быть выполнен только за счет фонда F_A . Если условие $Q_A(0) \geq Q(1, t)$ выполняется для лет $t = 1, \dots, \tau$, то для года τ – последнего года, в котором еще можно покрыть потребности в нефти за счет F_A , выполняется неравенство $Q(1, \tau) \leq Q_A(0) < Q(1, \tau+1)$. Дефицит нефти ΔQ_A в F_A , который надо в году $\tau+1$ покрыть за счет фонда F_C , равен $\Delta Q_A = Q(1, \tau+1) - Q_A(0)$. Тогда при $t \in [1, \tau]$ в фонде F_A притоки $F_A^+(t) = 0$, $F_A^+(\tau+1) = \Delta Q_A / K_A$, а при $t \in [\tau+2, T]$ $F_A^+(t) = Q(t) / K_A$. Итак, начиная с года $t = \tau+1$, фонду F_A нужна подпитка из фонда F_C . При этом, начиная с $t = \tau+2$, $F_A(t) = F_A(T)$, что означает работу «точно в срок». Ясно, что это оптимальная стратегия по критерию минимума приведенных капитальных затрат на освоение нефтеносных участков (без учета операционных расходов).

Пусть на интервале $[K_{AcM}, K_{AcM}]$ экспертом задано распределение вероятностей. В дополнение к K_A разыграем значение K_{Ac} и найдем величину $Q_C(0)$, отвечающую этому значению коэффициента перевода K_{Ac} . Начиная с года $t = \tau+1$, производственная программа может быть выполнена за счет привлечения ресурсов фонда F_C , если $Q_C(0) \geq Q(\tau+1, t)$. Если это условие выполняется для лет $t \in [\tau+1, \sigma]$, то для года σ – последнего года, в котором еще можно покрыть потребности в нефти за счет F_C , выполняется неравенство $Q(\tau+1, \sigma) \leq Q_C(0) < Q(\tau+1, \sigma+1)$. Дефицит нефти ΔQ_C в фонде F_C , который надо в году $\sigma+1$ покрыть за счет фонда F_D , равен $\Delta Q_C = Q(\tau+1, \sigma+1) - Q_C(0)$. Тогда при $t \in [1, \sigma]$ в фонде F_C притоки $F_C^+(t) = 0$, $F_C^+(\sigma+1) = \Delta Q_C / (K_A K_{Ac})$, а при $t \in [\sigma+2, T]$ $F_C^+(t) = Q(t) / (K_A K_{Ac})$. Итак, с $t = \sigma+1$ фонду F_C нужна подпитка из фонда F_D . При этом, начиная с года $t = \sigma+2$, и до конца периода прогнозирования $F_C(t) = F_C(T)$.

Аналогичным образом для фонда F_D имеем: $Q_D(0) = [F_D(0) - F_D(T)] / K_A K_{Ac} K_{Cd}$. Начиная с года $t = \sigma+1$, производственная программа $Q(t)$ может быть выполнена за счет привлечения ресурсов фонда F_D , если $Q_D(0) \geq Q(\sigma+1, t)$. Если это условие имеет место для лет $t \in [\sigma+1, \mu]$, то для года μ – последнего года, в котором еще можно покрыть потребности в нефти за счет F_D , выполняется неравенство $Q(\sigma+1, \mu) \leq Q_D(0) < Q(\sigma+1, \mu+1)$. Дефицит нефти ΔQ_D в фонде F_D , который надо в году $\mu+1$ покрыть за счет подпитки извне, равен $\Delta Q_D = Q(\sigma+1, \mu+1) - Q_D(0)$. Тогда при $t \in [1, \mu]$ в фонде F_D притоки $F_D^+(t) = 0$, $F_D^+(\mu+1) = \Delta Q_D / (K_A K_{Ac} K_{Cd})$, а при $t \in [\mu+2, T]$ $F_D^+(t) = Q(t) / (K_A K_{Ac} K_{Cd})$. Итак, с года $t = \mu+1$ фонду F_D нужна подпитка извне. При этом, начиная с года $t = \mu+2$, $F_D(t) = F_D^+(T)$.

Таким образом, разыгрывая значения K_A, K_{Ac}, K_{Cd} , для каждой истории Монте-Карло находим величину $t = \tau$ и, тем самым, временной диапазон $[1, \tau]$, в котором $F_A^+(t) = 0$, значение $F_A^+(\tau+1)$ и $F_A^+(t) = Q(t) / K_A$ для $t \in [\tau+2, T]$. Аналогичным образом в случае фонда F_C находится величина $t = \sigma > \tau$ и, тем самым, временной диапазон $[1, \sigma]$, в котором $F_C^+(t) = 0$, значение $F_C^+(\sigma+1)$ и $F_C^+(t) = Q(t) / (K_A K_{Ac})$ для $t \in [\sigma+2, T]$; в случае фонда F_D находится величина $t = \mu > \sigma$ и, тем самым, временной диапазон $[1, \mu]$, в котором $F_D^+(t) = 0$, $F_D^+(\mu+1)$ и $F_D^+(t) = Q(t) / (K_A K_{Ac} K_{Cd})$ для $t \in [\mu+2, T]$.

Таким образом, теперь не нужно, разыграв значения коэффициентов перевода, каждый раз решать оптимизационную задачу. В предложенном алгоритме сделано упрощение по сравнению с моделью (1 – 3), состоящее в отказе от ограничений на возможности пополнения фондов, которые связаны с производственным и финансовым потенциалом инвестора. Этот недостаток преодолевается за счет использования диалогового режима «эксперт – модель». Оперативность получения результатов позволяет эксперту быстро оценить реализуемость текущего решения модели и ввести необходимые коррективы в исходные данные. Такие модели были названы нами «экспертными моделями» [Пороскун и др., 2002].

Вернемся к ситуации а). Здесь программа $Q(t)$ может быть выполнена только за счет фонда $F_A(t)$. Мы не можем, однако, теперь гарантировать выход на запланированное значение $F_A(T)$: получаемая при каждом текущем значении K_A конечная величина фонда A составляет $F_A(T) = F_A(0) - Q(1, T) / K_A$. Эти величины лежат в интервале $[F_A(0) - Q(1, T) / K_{Am}, F_A(0) - Q(1, T) / K_{AM}]$, и на этом интервале существует рассчитываемое распределение вероятностей. Фонды F_C и F_D остаются в этом случае нетронутыми.

В варианте б) мы можем выйти на значение $F_A(T)$ и не можем гарантировать достижение $F_C(T)$, а в варианте в) выходим на $F_A(T)$ и $F_C(T)$, но не гарантируем значения $F_D(T)$.

Обратим внимание на еще одно обстоятельство. С первого взгляда кажется, что для формирования распределений вероятности для ненулевых пополнений каждого года периода прогнозирования необходимо разыгрывание значений коэффициентов перевода. Это не совсем так. Дело в том, что, как мы видели, начиная с некоторого момента времени, случайные величины $F^+(t)$, нормированные на $Q(t)$, для всех последующих лет имеют одинаковый закон распределения вероятностей (свой для каждого фонда), следующий законам для случайных величин $1/K_A$, $1/(K_A K_{Ac})$, $1/(K_A K_{Ac} K_{cd})$ соответственно. Это позволяет говорить о том, что величины $F^+(t)/Q(t)$ обладают свойством масштабной инвариантности, где $Q(t)$ играют роль масштабных коэффициентов, что упрощает расчеты результирующих распределений для $F^+(t)$. Эти распределения позволяют эксперту получить для каждого года прогнозного периода оценки для верхних границ пополнения фондов на выбранном им уровне вероятности.

Заключение

Предложен алгоритм расчета динамики объемов пополнения фондов нефти в их системе. Алгоритм отправляется от простой балансовой модели процесса пополнения. Результаты расчетов эквивалентны решениям задачи в оптимизационной постановке. Использование алгоритма позволяет упростить расчеты в условиях неопределенности, обходя необходимость многократного решения оптимизационной задачи. Дополнительным упрощающим расчет обстоятельством служит свойство масштабной инвариантности для распределений вероятностей компонентов пополнения фондов нефти.

Благодарности

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Библиография

- [Кноринг, 1992] Кноринг Л. Стратегия подготовки запасов нефти и газа. СПб: Недра, 1992.
- [Пороскун и др., 2002] Пороскун В., Стернин М., Шепелев Г. Экспертные модели прогнозных оценок подготовки и освоения запасов нефти. // Искусственный интеллект, 2002, №2, сс. 540-546.
- [Стернин и др., 2007] Стернин М., Шепелев Г., Шепелев Н. Интеллектуальная поддержка прогнозных экспертных оценок подготовки и освоения запасов нефти. // «Интеллектуальные системы», Труды международной научно-технической конференции AIS'07. М.: Физматлит 2007, т. 2, сс. 105 – 106.

Информация об авторах



Михаил Стернин – старший научный сотрудник Института системного анализа РАН.
Россия, 117312, Москва, просп. 60-летия Октября, ИСА РАН; e-mail: mister@isa.ru



Геннадий Шепелёв – заведующий лабораторией Института системного анализа РАН.
Россия, 117312, Москва, просп. 60-летия Октября, ИСА РАН; e-mail: gis@isa.ru

РЕШЕНИЕ ЗАДАЧИ АВТОМАТИЧЕСКОЙ СЕПАРАЦИИ ФУЛЛЕРЕНСОДЕРЖАЩЕЙ СУСПЕНЗИИ

А.В. Тимофеев, П.Г. Иванов

Abstract: Синтез фуллереноподобных структур в электрической дуге по своей сути является процессом неизбежного характера, т.е. в данный момент нет достаточной информации и технологии, позволяющей получать фуллерены одного вида при возгонке графитовых стержней. Вследствие данной особенности, становится актуальной задача разделения растворов фуллереноподобных структур [Мищенко, Ткачев, 2008].

На сегодняшний день качественный и количественный анализ получаемых экстрактов фуллеренов можно произвести только на комплексе спектральных установок. Такой тип анализа синтезируемых фуллеренов является неприменимым для его внедрения в систему управления процессами из-за ограничений связанных со временем проведения комплекса анализов, а так же требованиями подготовки проб.

В большинстве случаев из-за отсутствия специализированных сенсоров, встраиваемых непосредственно в процесс, качественный состав выходного потока промышленных установок анализируется человеческим глазом или на основе эмпирических данных, полученных на основе приблизительного времени выхода индивидуальных фракций. Оба метода не дают высокой степени чистоты получаемых растворов. Таким образом, для решения задачи автоматической сепарации необходим поиск параметров, измерение которых позволит с высокой точностью разделять поток на мономолекулярные фракции.

Keywords: сенсоры, информационное моделирование, фуллерены, фотометрический анализ

Постановка задачи

Вычисление концентрации в многокомпонентной смеси является не тривиальной задачей. Дополнительную сложность создает природа самих фуллеренов, так как все они представляют собой молекулы, состоящие из атомов углерода, которые отличаются только строением и количеством атомов. Поэтому смеси фуллеренов дают очень схожие электронные спектры.

Спектрофотометрия входит в комплекс лабораторных методов, позволяющих однозначно определить качественный и количественный состав растворов, содержащих фуллерены. Поэтому решение задачи предполагается основывать на этом методе анализа после этапа жидкостной хроматографии. Выбор такой структуры системы позволяет анализировать динамическое изменение концентраций и позволит разделять смесь на фракции с чистотой, превышающей существующие промышленные методы сепарации.

В работе М.И.Булатова и И.П.Калинкина [Булатов, Калинин, 1986] оптическая плотность любой среды, содержащая ограниченное число компонентов, не взаимодействующих химически друг с другом, описана как суперпозиция оптических плотностей компонентов смеси при той же длине волны:

$$A_{\lambda} = a_{1\lambda} + a_{2\lambda} + a_{3\lambda} + \dots + a_{n\lambda} \quad (1)$$

причем каждая из парциальных плотностей равна:

$$a_{i\lambda} = \varepsilon_{i\lambda} \cdot C_i \cdot l \quad (2)$$

здесь $\varepsilon_{i\lambda}$ -молярный коэффициент светопоглощения вещества i при длине волны λ ; l - толщина поглощающего слоя, (см); C_i - концентрация поглощающего вещества, (моль/л).

В связи с особенностью процесса экстракции и разделения первоначального раствора смеси фуллеренов в потоке, выходящем из колонны хроматографа концентрация двух видов фуллеренов C60 и C70, достигает 98% [Попов, 2007].

Таким образом, можно уточнить задачу следующим образом: необходимо построить систему автоматической сепарации 2-х компонентной смеси на три фракции, на растворы чистого C60, смесь C60+C70 и чистого C70. Вторая фракция является промежуточной и будет присутствовать всегда.

Для реализации спектрофотометрического анализа необходимо исследование системы линейных уравнений:

$$\begin{cases} A_{\lambda 1} = \varepsilon_{1\lambda 1} \cdot C_1 \cdot l + \varepsilon_{2\lambda 1} \cdot C_2 \cdot l \\ A_{\lambda 2} = \varepsilon_{1\lambda 2} \cdot C_1 \cdot l + \varepsilon_{2\lambda 2} \cdot C_2 \cdot l \end{cases} \quad (3)$$

Но, как отмечалось ранее, на данный момент не существует двухволновых спектрофотометров работающих в условиях технологического потока.

Сенсор

Для физической реализации процесса автоматической сепарации потребовался сенсор, способный вычислять концентрацию растворенных в смеси фуллеренов и сигнализировать момент начала выхода очередной фазы из хроматографа. Его реализация была выполнена на светодиодной базе с принципом работы, основанным на светопоглощении. Конструкция сенсора представляет собой проточную кювету, помещенная в собирающийся корпус с оптопарами, количество которых зависит от требуемых длин волн. В контексте данной задачи используется две пары светодиод-фотодиод, как чувствительный элемент и микроконтроллер для выполнения расчетов и анализа. Макет данного датчика представлен на рисунке 1.

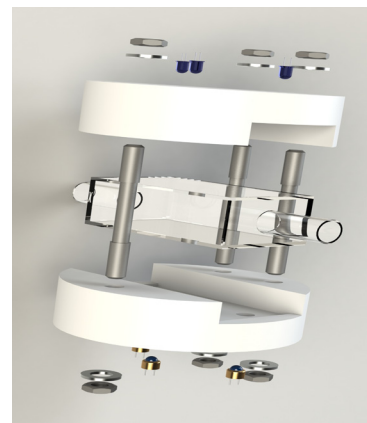


Рисунок 1. Прототип сенсора системы автоматической сепарации фуллеренов

В качестве математического аппарата для вычисления концентрации компонентов смеси применяется решение системы линейных уравнений (3). Для определения качественного состава потока используется отношение светопоглощения на двух различных длинах волн:

$$\frac{A_{\lambda 1}}{A_{\lambda 2}} = \frac{\varepsilon_{1\lambda 1}}{\varepsilon_{1\lambda 2}} + \frac{\varepsilon_{2\lambda 1}}{\varepsilon_{2\lambda 2}} \quad (4)$$

Уравнение (4) не зависит от количественной характеристики вещества, а определяется лишь физическими характеристиками компонентов. Поэтому изменение отношения будет наблюдаться только в моменты качественного изменения раствора проходящего через кювету. Как оказалось, область применения созданной системы может выходить за рамки установок по производству фуллеренов.

Другие области применения сенсора

В настоящее время в виду своих уникальных свойств фуллерены становятся все более популярными не только в лабораторных исследованиях, но и находят новое применения в существующих технологических процессах и при создании новых. Однако вследствие высокой цены, связанной с высокими затратами и потерями при выделении, применимость фуллеренов ограничена. Зачастую вместо чистых фуллеренов в производстве используют фуллеренсодержащую сажу, концентрация которых в ней достигает 15%

(производство каучуков, полимерных пленок и т.д.), а так же сажу, прошедшую этап экстракции при производстве фуллеренов. Формально эта сажа для производства фуллеренов является отходами, но может применяться как сырье в других сферах (строительные материалы и т.д.). В связи с этим требуется анализ, который обнаруживает присутствие фуллеренов в саже и который может дать количественную оценку их концентрации. Очень важно, что данные анализа можно подтвердить лабораторными приборами. При этом стоимость измерения и контроля качества будет значительно ниже и не будет требовать специально подготовленных для этого кадров.

Заключение

Важным результатом проделанной работы стало обнаружение информационного канала, по которому в процессе разделения хроматографического потока, удалось осуществить автоматическое управление. Экспериментальные результаты внедрения описанной системы автоматической сепарации углеродных наноматериалов показали, что своевременное автоматическое перенаправление потока, выходящего из колонны хроматографа, может обеспечить чистоту производимых фуллеренов не менее чем 98%. Важно отметить, что создание системы автоматического управления и интеллектуального контроля позволит полностью исключить человека оператора на описанном этапе производства.

Литература

[Мищенко, Ткачев, 2008] С.В. Мищенко, А.Г. Ткачев. Углеродные наноматериалы. Производство, Свойства, Применение. Москва. Машиностроение. 2008

[Булатов, Калинин, 1986] Булатов М.И., Калинин И.П. Практическое руководство по фотометрическим методам анализа. Ленинград. Химия. 1986

[Попов, 2007] Диссертация. Попов А.М. Свойства углеродных наноструктур и наноэлектромеханические системы на их основе. Троицк. 2007

Информация об авторах

Тимофеев Адиль Васильевич – заведующий лабораторией информационных технологий в управлении и робототехнике Санкт-Петербургского института информатики и автоматизации РАН, д.т.н., профессор, Заслуженный деятель науки РФ, 199178, Россия, Санкт-Петербург, 14-ая линия, д.39, СПИИРАН, tav@ias.spb.su

Иванов Павел Геннадьевич – аспирант Санкт-Петербургского института информатики и автоматизации РАН, waver_tech@mail.ru

МЕТОДЫ И АЛГОРИТМЫ ОБНАРУЖЕНИЯ ОБЪЕКТОВ ДЛЯ СИСТЕМЫ УПРАВЛЕНИЯ ПРОМЫШЛЕННОЙ БЕЗОПАСНОСТЬЮ В РАМКАХ КОНЦЕПЦИИ СОЗДАНИЯ КОМПЛЕКСНОЙ АСУ ГАЗОТРАНСПОРТНОЙ КОМПАНИЕЙ

Белоус Н.В., Борисенко В.П., Пономарев Ю.В., Борисенко И.В.

Abstract: Одним из перспективных подходов к решению проблемы "островной" автоматизации предприятий газотранспортной системы (ГТС) является создание комплексной автоматизированной системы управления (КАСУ), имеющей гибкую многоуровневую и распределенную архитектуру. В работе рассмотрены основные положения концепции создания КАСУ газотранспортной компании, в рамках которой предложен комплексный подход к построению автоматизированной системы управления промышленной безопасностью ГТС, объединяющий современные технологии системной интеграции корпоративной информации на основе единой инструментальной технологической платформы, а также методы обработки и анализа данных о потенциально опасных объектах с целью их своевременного обнаружения и поддержки принятия решений о применении своевременных превентивных мер. Оригинальной особенностью предлагаемого подхода является ориентация на методы сбора, формализованного представления и компьютерной обработки специализированной графической информации, в виде космических и аэрофотоизображений, а также данных реального видео, получаемых от разнообразных промышленных систем видеонаблюдения, которые в настоящее время получают все более широкое распространение для обеспечения надежной и безопасной эксплуатации объектов газотранспортной отрасли.

ACM Classification Keywords: I.3.5 Computational Geometry and Object Modeling

Введение

Для крупномасштабных компаний топливно-энергетического комплекса весьма перспективным подходом является комплексная "сквозная" автоматизация основных технологических и бизнес-процессов на основе разработки и внедрения комплексной автоматизированной системы управления (КАСУ) газотранспортной компанией (ГТК). В рамках концепции создания КАСУ [Борисенко, 2003],[Борисенко, 2004],[Борисенко, 2008] было всесторонне обоснована актуальность построения современных компьютерных систем управления промышленной безопасностью (ПБ) в состав которых входят относительно новые и весьма перспективные для ГТК подсистемы мониторинга потенциально опасных объектов на базе современных аэрокосмических и видеоинформационных технологий.

В статье рассматриваются гибридные методы и алгоритмические средства компьютерной обработки изображений, предназначенные для раннего обнаружения потенциально опасных объектов, предотвращения аварий и прогнозирования чрезвычайных ситуаций. Предложенные методы и средства предварительно были исследованы в имитационном режиме на основе организационно-технологических моделей базовых элементов газотранспортных систем, а также прошли достаточно широкую практическую апробацию.

Опыт разработки и внедрения ИАСУ ДК "Укртрансгаз"

Прообразом КАСУ ГТК был проект создания интегрированной автоматизированной системы управления (ИАСУ) ДК «Укртрансгаз», реализация которого был одним из первых шагов на пути построения

полнофункциональной комплексной автоматизированной системы управления компанией, важным с точки зрения накопленного опыта [Борисенко, 2004], [Борисенко, 2008]. ИАСУ ДК "Укртрансгаз" создавалась, начиная с 1999 г., специализированным отраслевым институтом как *заказная система*, что позволило обеспечить более точное соответствие разработок требованиям Заказчика с учетом их изменений и менее затратную итеративную (поэтапную) технологию внедрения. В настоящий момент система внедрена на нескольких предприятиях газотранспортной отрасли Украины.

Анализ практических результатов и недостатков проектирования ИАСУ ДК "Укртрансгаз" показал, что использование заказного системотехнического подхода к созданию КАСУ ГТК в современных условиях является недостаточным. Для автоматизации унифицированных видов деятельности, типовых бизнес-процессов отраслевого характера наиболее перспективным является использование покупных параметризованных программно-информационных комплексов.

Поэтому в дальнейшем в качестве базового методологического принципа разработки КАСУ ГТК предлагается совместное использование двух основных системотехнических подходов - 1) построение унифицированного ядра КАСУ на основе приобретения и внедрения покупной тиражируемой системы; 2) самостоятельная разработка специализированных подсистем и компонент КАСУ как заказных систем.

При этом эффективная реализация объединения данных подходов базируется на использовании специализированного системного инструментария в виде корпоративной интеграционной технологической платформы [Борисенко, 2007]. На основе проведенных исследований в состав интеграционной платформы было предложено включить такой основной набор сервисных инструментальных средств:

- ведения единого репозитория метаданных общесистемного уровня;
- поддержка объектно-ориентированной технологии интеграции данных;
- поддержка технологий веб-интеграции и создания веб-порталов;
- поддержка технологий хранилищ данных и OLAP;
- поддержка сервис-ориентированной архитектуры и технологий интеграции;
- поддержка технологии обмена сообщениями.

Базовые положения концепции создания КАСУ ГТК

Главной целью разработки и внедрения корпоративной автоматизированной системы управления ГТК является целостное, "сквозное", взаимосогласованное решение всех основных задач управления технологическими, техническими и организационными объектами компании.

К классу КАСУ ГТК относятся интегрированные компьютеризированные системы управления, предназначенные для комплексной автоматизации всех основных видов деятельности и бизнес-процессов в компании.

В КАСУ реализован процесс объединения в единую систему управления самостоятельно функционирующих автоматизированных систем на основе регулярного и упорядоченного обмена информацией между ними.

Все автоматизированные системы, включаемые в состав КАСУ должны удовлетворять единым общесистемным архитектурным требованиям.

Основной системотехнический эффект от внедрения КАСУ ГТК заключается в переходе от локализованного, слабо синхронизированного, разнородного набора автономных подсистем к единой, комплексной, системной архитектуре, которая строится на основе универсального набора современных сервисных инструментальных средств поддержки современных технологий интеграции – корпоративной интеграционной технологической платформы (КИТП) [Борисенко, 2007].

Разработка и внедрение КАСУ ГТК обеспечивает следующие основные бизнес-преимущества:

- повышение уровня стандартизации и унификации элементов системы, что существенно увеличивает ее гибкость и адаптивность к изменениям внешней среды;
- создание единого информационного пространства, обеспечивающее существенное снижение дублирования, несогласованности и противоречивости данных, а также повышение достоверности информации, предоставляемой конечным пользователям КАСУ ГТК;
- общее снижение уровня рисков и суммарных затрат на создание, внедрение и владение системой;
- в конечном итоге успешное внедрение КАСУ оказывает существенное влияние на повышение качества, надежности и эффективности функционирования газотранспортной системы в целом.

В то же время, разработка и внедрение КАСУ ГТК является сложным, многоэтапным и весьма трудоемким процессом, осуществляемым в условиях жестких ограничений по времени и ресурсам.

Система управления промышленной безопасностью ГТС

В настоящее время в число наиболее значимых и актуальных проблем газотранспортной отрасли вошла разработка на новой технологической платформе современной автоматизированной системы управления промышленной безопасностью (АСУ ПБ). АСУ ПБ включает как традиционные подсистемы охраны труда безопасности, пожарной безопасности, экологической безопасности, так и оригинальные относительно новые и весьма перспективные для ГТП подсистемы и комплексы задач:

- мониторинга, обнаружения и оповещения о потенциально опасных объектах в охранной зоне линейной части магистральных газопроводов, а также на территории крупных организационно-технологических объектов - компрессорных и газораспределительных станций, подземных хранилищ газа;
- сбора данных, обработки, анализа данных и прогнозирования чрезвычайных ситуаций (ЧС) природного и техногенного характера.

Важность разработки АСУ ПБ связана с рядом свойств и особенностей газотранспортной отрасли:

- старение газотранспортной сети (в настоящее время до 80% газопроводов срок эксплуатации превышает 20-25 лет, что существенно увеличивает риск возникновения аварий и чрезвычайных ситуаций);
- появление принципиально новых опасностей и угроз, связанных с строительством морских трубопроводов, а также их прокладка в горной местности, где возможны селевые потоки, снежные лавины, оползни и т.п.;
- возможность появления ЧС, связанных с халатностью эксплуатационного персонала и угрозами терроризма.

Создание автоматизированных систем мониторинга потенциально опасных объектов, анализа и прогнозирования ЧС, лежащих в основе АСУ ПБ, с использованием современных аэрокосмических методов является новым и весьма перспективным подходом к комплексной автоматизации системы безопасности. Одним из наиболее информативных и технологичных источников данных для таких систем является комплексный дистанционный аэрокосмический мониторинг объектов ГТП, позволяющий получать снимки наблюдаемой территории, как при помощи космических аппаратов (спутников), так и с использованием средств современной пилотной и беспилотной авиации. Аэрокосмические снимки можно использовать как непосредственно, так и в комплексе с другими данными для автоматизированного решения следующих прикладных задач:

- оценка целостности и безопасности объектов ГТК, а также безопасности окружающей их среды;
- выявление и картографирование зон вероятного возникновения техногенных и природных аварий и катастроф;
- обнаружение и мониторинг экологических негативных процессов, протекающих в охранной зоне газопроводов, связанных с посторонними источниками загрязнения и техногенного влияния;

– выявление и картографирование изменений в геодинамически активных зонах, связанных с территориями подземных газохранилищ, на основе радарной съемки и интерферометрии, раннее обнаружение сдвиговых поверхностных деформаций в зонах расположения трубопроводов;

– мониторинг территорий в охранных зонах газотранспортного предприятия на предмет наличия зон самозахвата и участков несанкционированного строительства жилых и промышленных объектов.

Кроме этого, мониторинг земной поверхности позволит своевременно предсказывать природные катастрофы, например, по анализу формы и размеров грозových облаков и дымовых завес во время пожаров.

Решение таких задач связано с анализом объектов на аэро-, фото-изображениях, с последующим их распознаванием и идентификацией.

Методы обнаружения объектов

Множество известных методов обнаружения объектов можно разделить на 2 основных класса: 1) обнаружение объектов с частично или полностью размытой границей [Белоус, 2009] и 2) обнаружение объектов, имеющих четко заданную границу [Biliou, 2008].

Обнаружение объектов с размытой границей

К такому классу объектов можно отнести изображения облаков на снимках или дымовые облака пожаров, которые необходимо выделить.

Исходное изображение аэрокосмической съемки имеет большое количество мелких неинформативных объектов: деревья, люди, наземный транспорт, мелкие постройки и т.д. Применение сглаживающей фильтрации по Гауссиану на первом этапе дает возможность снизить влияние неинформативных объектов на исходное изображение, что приводит к сглаживанию и при этом размытию границы облаков.

На втором этапе для восстановления границы облаков, а также для повышения ее четкости выполняется контрастоповышающая модификация изображения. Данный прием позволяет более точно выделить светлые границы облаков и дыма.

Третий этап выполняет градирование изображения для разделения изображения на отдельные зоны. Для максимальной точности применяется градирование по алгоритму Собеля [Белоус, 2009], в котором используется восемь отсчетов яркости в окрестностях центральной точки:

$$G(x, y) = \sqrt{Gx_{x,y}^2 + Gy_{x,y}^2},$$

$$G(x, y) \cong |Gx_{x,y} + Gy_{x,y}|,$$

$$Gx_{x,y} = [A_{x-1,y-1} + 2A_{x-1,y} + A_{x-1,y+1}] - [A_{x+1,y-1} + 2A_{x+1,y} + A_{x+1,y+1}],$$

$$Gy_{x,y} = [A_{x-1,y-1} + 2A_{x,y-1} + A_{x+1,y-1}] - [A_{x-1,y+1} + 2A_{x,y+1} + A_{x+1,y+1}].$$

где $G(x, y)$ – значения яркости точки в координатах x, y после проведения операции градиента;

$Gx_{x,y}$ – промежуточное значение градиента по параллелям x ;

$Gy_{x,y}$ – промежуточное значение градиента по параллелям y ;

$A_{x,y}$ – значения яркости точки в координатах x, y .

На четвертом этапе выполняется объединение изображения в зоны сходной яркости с использованием алгоритма «Волшебная палочка». Для работы данного метода указывается одна или несколько начальных точек объекта, а алгоритм выделяет окрестные пиксели с похожим цветом и добавляет

выделенную область к объекту. Для оценки «похожести» задается матрица порогов чувствительности расстояния между цветами. Чем выше порог чувствительности, тем большая зона пикселей выделится на одном шаге. При этом в область «похожих» цветов выделяются только связанные пиксели. Метод эффективно работает при выделении достаточно монотонных по цвету объектов. При сильных вариациях цвета аккуратно отделить объект от фона с помощью данного алгоритма невозможно.

Последний, завершающий этап, разделяет изображение на зоны высокой яркости (облака, дым), и зоны низкой яркости (территория земной поверхности).

Обнаружение объектов с четкой границей

Разработаны методы обнаружения объектов по определенным характерным признакам и с четко заданной границей. К таким классам объектов могут быть отнесены всевозможные строительные сооружения и транспортные средства, нахождение которых в охранных зонах газопровода может привести к его повреждению.

Основная проблема, сопряженная с использованием методов данной группы является отсутствие временных данных, позволяющих однозначно определить появление новых объектов, а также наличие и изменение положения теней, отбрасываемых объектами, что усложняет введение устойчивых признаков.

Предложены методы обнаружения появившихся или движущихся объектов по результатам анализа изменений на пространственно-временных аэрофотографических и космических снимках. Обнаружение объектов осуществляется путем формирования карты неизменных объектов и земной поверхности, т.е. фона, после чего появление объектов обнаруживается разностным сравнением новых снимков. Основным недостатком этого подхода также является влияние изменения теней и освещенности земной поверхности, растительности и объектов, что, в случае анализа спутниковых снимков, ограничивает применение пространственно-временного подхода к обнаружению глобальных изменений (растительность полей, обнаружение затоплений, пожаров), а также крупных строительных объектов.

Однако развитие методологии моделирования, обнаружения и удаления теней на аэрофотоснимках снимках позволяет после дополнительной обработки проводить обнаружение и небольших транспортных объектов. Методы обнаружения транспортных объектов позволяют определять размеры и скорость объектов, а также осуществлять слежение, но требуют высокой частоты обновления снимков и высокого разрешения, что не всегда возможно в области мониторинга аварийно опасных зон из-за необходимости охвата больших по площади территорий и невозможно для спутникового мониторинга.

Еще одним эффективным подходом обнаружения объектов, как вновь появившихся, так старых при картографировании, является применение фотограмметрии для восстановления трехмерных координат земной поверхности. Восстановление трехмерных данных осуществляется благодаря линейному прохождению спутника или пролета летательных аппаратов над обследуемой местностью и перекрытию соседних снимков, что позволяет использовать пары/тройки соседних снимков для нахождения соответствий и восстановления координат точек поверхности. При этом отделение теней значительно упрощается, появляются новые данные для анализа не только поверхностных, но и высотных размеров объектов. Проблемой подхода является низкая скорость восстановления трехмерных данных, что исключает использование подхода в условиях ограничений по времени при мониторинге аварийных объектов.

Одним из наиболее эффективных подходов для снижения временных затрат при восстановлении трехмерных данных является уменьшение количества сопоставляемых точек за счет выделения некоторого множества характерных точек. Данный подход используется преимущественно для стереометрического восстановления поверхностей, когда пары изображений представлены в различных масштабах или подвержены различным проективным преобразованиям. Его применение для восстановления трехмерных данных по горизонтально сдвинутой паре изображений также возможно и может снизить вычислительные затраты на сопоставление точек и вычисление трехмерных координат.

Предлагается применение комбинированного подхода, использующего пространственно-временное представление восстановленных трехмерных данных и масштабно-пространственных представлений снимков при восстановлении трехмерных данных для значительного снижения количества анализируемых точек путем выделения наиболее информативных граничных областей.

Задача восстановления трехмерных данных по паре аэрокосмических снимков является более простой в вычислительном плане по сравнению с подобными задачами из других областей, где необходимо учитывать не только горизонтальное смещение точек при изменении точки наблюдения, но и вертикальные смещения, обусловленные перспективными искажениями. Т.е. задача восстановления карты поверхности земли в трехмерных координатах сводится к нахождению соответствующих точек на снимках, сделанных с различных точек наблюдения и вычислению третьей координаты по относительному смещению таких точек [Bilious, 2011].

Предлагаемое применение масштабно-пространственного представления снимков для снижения числа точек к сопоставлению предполагает последовательное вычисление функции кривизны в строках изображений при их сканировании в любом порядке на различных уровнях масштаба. При этом, представляя интенсивность функцией сдвига по горизонтали, получаем непрерывную кривую, заданную уравнением в явной форме $y = y(x)$, для которой кривизна вычисляется по формуле [Linderberg, 1994]:

$$k = \ddot{y} / (1 + \dot{y}^2)^{3/2}.$$

При работе с растровыми изображениями спутниковых снимков получаем только дискретную кривую. Такая дискретная кривая в общем случае может содержать разрывы и отдельно стоящие, "выбитые" точки. Как дифференциальная характеристика, кривизна является очень неустойчивой к дискретизации и зашумлению изображения и не может быть использована для нахождения характерных точек применительно к исходной кривой. Именно поэтому предлагается представление кривизны на различных уровнях масштаба, для чего используется свертка функции кривизны Гауссовым ядром увеличивающегося размера [Linderberg, 1994]:

$$L(x, \sigma) = \int_{\xi = -\infty}^{+\infty} g(\xi, \sigma) f(x - \xi) d\xi,$$

$$g(x; \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-x^2 / 2\sigma^2},$$

где $g : \mathbb{R} \times \mathbb{R}_+ \setminus \{0\} \rightarrow \mathbb{R}$ является Гауссовым ядром.

В результате отслеживания точек перехода кривизны через ноль, отсеиваются все точки, не представляющие сильных изменений по интенсивности на исходном снимке, а при наложении определенных ограничений к уровню масштаба остается множество характерных точек, влияние которых на перепады интенсивности строк снимков значительно даже на высоких уровнях свертки. Такие точки гарантировано присутствуют на двух изображениях одной и той же местности и, в большинстве случаев, представляют границы определенных объектов, даже если точка зрения изменяется. Т.е. опираясь на характерные точки, можно восстановить основные высотные изменения поверхности, выделить объекты, расположенные на земной поверхности.

Сравнивая поверхности определенной территории, восстановленные в различные моменты времени, можно по расхождению интенсивности и геометрической (высота над средним уровнем земной поверхностью) разнице с высокой точностью выявить появление или исчезновение объектов размера, зависящего только от разрешения снимков.

Заклучение

В работе была обоснована актуальность, рассмотрены принципиальные основы, базовые подходы и основные стратегии создания комплексной автоматизированной системы управления ГТК.

Анализ опыта создания интегрированной АСУ Компании "Укртрансгаз", как прообраза полнофункциональной КАСУ ГТК, позволил обосновать применение оригинального набора технологий интеграции на основе совместного использования: 1) "заказного" подхода для разработки высокоэффективных специализированных автоматизированных систем и подсистем; 2) подхода, базирующегося на приобретении параметризованных покупных коммерческих систем для автоматизации технологических и деловых процессов межотраслевого уровня унификации; 3) корпоративной инструментальной интеграционной технологической платформы в качестве единой технологической основы построения высокоэффективной системной архитектуры.

В работе были рассмотрены методы и модели компьютерной обработки аэрокосмических данных, лежащие в основе одной из наиболее важных и перспективных подсистем в составе КАСУ ГТК – комплексной АСУ промышленной безопасности. Для обеспечения эффективного применения этих технологий в составе АСУ ПБ был проведен анализ, рассмотрена классификация методов обнаружения значимых объектов ГТС, рассмотрены модели и алгоритмы их реализации, рассмотрены особенности и оригинальная реализация предложенного подхода к анализу и распознаванию аэрокосмических изображений.

Библиография

- [Борисенко, 2003] Борисенко В.П., Принципы построения и методология разработки интегрированной автоматизированной системы управления газотранспортной компанией. // АСУ и приборы автоматики. 2003. – Вып. 125. – С. 86–91.
- [Борисенко, 2004] Борисенко В.П., Колодяжный В.В. Пономарев Ю.В. Методология построения и опыт реализации распределенной интегрированной системы управления газотранспортной системой Украины. // Трубопроводные системы энергетики. Управление развитием и функционированием.- Новосибирск: Наука, 2004.-461 с.
- [Борисенко, 2008] Борисенко В.П., Медведева Л.М., Борисенко Т.І. Базові положення концепції створення комплексної автоматизованої системи керування основними бізнес-процесами підприємств нафтогазової галузі. // АСУ и приборы автоматики-2008.- № 144.- с. 88-96.
- [Борисенко, 2007] Борисенко В.П., Скляр С.А., Чмелев Ю.Н., Охрименко С.С., Метод построения и технология адаптивной синхронизации структурно-инвариантного объектно-ориентированного ядра распределенной базы данных ИАСУ ДК "Укртрансгаз"// Проблемы нефтегазовой промышленности: Сб. научн. трудов. Вып.5. – Киев, 2007. – с. 486-492.
- [Белоус, 2009] Белоус Н., Борисенко В., Левыкин В., Макивский Д., Зайцева А. Метод выделения значимых данных на изображениях изохромных линий для систем бесконтактного измерения внутриглазного давления// International Book Series «Information Science and Computing» №8. Classification, Forecasting, Data Mining // Supplement to International Journal «Information Technologies and Knowledge». ITHEA, SOFIA, 2009 – p.148-154.
- [Bilious, 2008] Bilious N., Kobzar G. Automatic ECG Analysis for Preliminary and Detailed Diagnostics Based on Scale-space Representation // International Book Series «Information Science and Computing» №7 Artificial Intelligence and Decision Making // Supplement to International Journal «Information Technologies and Knowledge». ITHEA, SOFIA, 2008 – p. 53-60- A4.
- [Bilious, 2011] Bilious N., Borisenko V. and ect. Intelligent Data Processing in Global Monitoring for Environment and Security//ITHEA, Kiev-Sofia, 2011, First Edition. - ISBN: 978-954-16-0045-0.
- [Linderberg, 1994] Linderberg T. Scale Space Theory in Computer Vision, Kluwer Academic Publishers. - 1994, 440p.

Информация об авторах



Белоус Наталья Валентиновна – заведующий лабораторией "Информационные технологии в системах обучения и машинного зрения", к.т.н., профессор каф. ПО ЭВМ Харьковского национального университета радиоэлектроники. Харьков. Украина. e-mail: belous@kture.kharkov.ua

Распознавание образов и изображений, обработка потокового видео, системы обучения и тестирования знаний.



Борисенко Виктор Петрович – к.т.н. доцент, член-корреспондент украинской нефтегазовой академии наук, ученый секретарь Совета Научно-исследовательского и проектного института транспорта газа, начальник отдела информационных технологий. Украина, 61004, г. Харьков, ул. Маршала Конева, 16, e-mail: vborisenko@itransgaz.com

Методологии проектирования и разработки корпоративных информационных систем, моделирование сложных систем, технологии разработки знание - базированных систем и программирования, методы и средства создания объектно-реляционных баз данных.



Пономарёв Юрий Владимирович – к.т.н. доцент, академик украинской нефтегазовой академии наук, заместитель директора по научной работе Научно-исследовательского и проектного института транспорта газа, Украина, 61004, г. Харьков, ул. Маршала Конева, 16, e-mail: ponom@itransgaz.com

Современные информационные технологии, моделирование сложных систем, методологии программирования, математическое моделирование, теоретическая расходомерия и метрология



Борисенко Илья Викторович – студент, Харьковский национальный университет радиоэлектроники, г. Харьков, Украина, e-mail: iliich1991@mail.ru

Компьютерная инженерия, web-базированная разработка, современные технологии проектирования и программирования информационных систем

РАСПОЗНАВАНИЕ ЗДАНИЙ НА СПУТНИКОВЫХ СНИМКАХ СВЕРХВЫСОКОГО РАЗРЕШЕНИЯ

Светлана Дьяконова

Аннотация. В работе рассматривается применение нечеткого классификатора NEFClass к задаче распознавания зданий на спутниковых снимках сверхвысокого разрешения. Проводится сравнительный анализ методов обучения нейронной сети: генетического, градиентного и метода сопряженных градиентов касательно задачи распознавания зданий с различными формами крыши.

Ключевые слова: распознавание зданий на спутниковых изображениях, нечеткий классификатор NEFClass, методы обучения нейронных сетей.

ACM Classification Keywords: 1.4.7 Computing Methodologies - IMAGE PROCESSING AND COMPUTER VISION – Feature Measurement - Size and shape. G.1.6 Mathematics of Computing – NUMERICAL ANALYSIS – Optimization- Gradient methods. I.4.8 Computing Methodologies - IMAGE PROCESSING AND COMPUTER VISION – Scene Analysis - Object recognition. I.5.1 Computing Methodologies - PATTERN RECOGNITION - Models - Neural nets.

Введение

Задача распознавания зданий на спутниковых изображениях сверхвысокого разрешения является важной частью в решение задачи автоматической интерпретации данных получаемых из систем дистанционного зондирования земли. Применение такой системы распознавания зданий возможно, например, для автоматического анализа использования земельных ресурсов, для поддержания городских геодезических данных в актуальном состоянии, в частности для контроля качества и автоматического обновления геодезических данных. Предложенные ранее работы по теме распознавания зданий разработаны с учетом того, что исследования проводятся на спутниковых снимках, взятых из одного региона, где все здания имеют определенный оттенок и форму крыши. Описываемый в данной статье подход позволяет использовать систему распознавания для различных видов населенных пунктов, это может быть город, пригород или населенный пункт в пустынном регионе. В частности в работе рассматривается три вида крыши зданий: плоская, круглая и треугольная крыша. Каждая из этих форм соответствует одному из типов построек. Здания с плоскими крышами преобладают в городских территориях, и соответствуют высокоэтажному строению, треугольные крыши наиболее часто встречаются в домах загородного типа, круглые крыши зданий преобладают в населенных пунктах находящихся в пустынных регионах. На основе знаний о геометрической форме крыши зданий были подобраны оптимальные признаки для распознавания. В качестве классификатора была выбрана архитектура системы NEFClass, которая впоследствии была программно реализована с тремя алгоритмами обучения: генетическим, градиентным и сопряженных градиентов.

Описание архитектуры нечеткого классификатора NEFClass

Модель NEFClass [Зайченко, 2008] используется для определения класса или категории полученных входных образцов (паттернов). Паттерны — это векторы признаков $x = (x_1, \dots, x_n) \in R^n$, а класс C — подмножество R^n . Принимаем, что пересечение двух разных классов является пустым. Величина признака паттерна представляется нечетким множеством, и классификация определяется множеством лингвистических правил. Для каждого входного признака x_i существует q_i нечетких множеств $\mu_1^i, \dots, \mu_{q_i}^i$. Также есть база правил, которая вмещает k нечетких лингвистических правил R_1, \dots, R_k .

База правил представляет собой аппроксимацию функции (неизвестной) $\varphi: R^n \rightarrow \{0,1\}^m$, которая представляет задачу классификации, где $\varphi(x) = (c_1, \dots, c_m)$ такая, что $c_i = 1, c_j = 0$ ($j \in \{1, \dots, m\}, j \neq i$). Значит, x принадлежит классу $C_i \rightarrow [0,1]^m$.

Обозначим наибольший компонент каждого вектора c единицей ($c = 1$), а все другие компоненты положим равными нулю.

Нечеткие множества и лингвистические правила, обеспечивающие выполнение такой аппроксимации и определяющие результирующую систему NEFClass, будут получены из множества примеров для обучения. На Рис. 1 приведена система NEFClass, которая классифицирует входные образцы с двумя признаками и двумя отдельными классами, используя пять лингвистических правил.

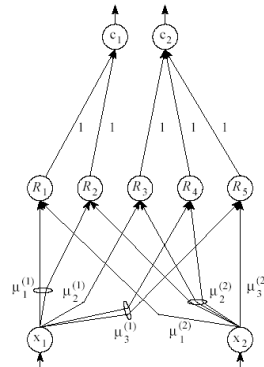


Рис.1 Пример нейросети NEFClass с 2 признаками, 5 правилами и 2 классами

NEFClass — это трехслойный нечеткий перцептрон со следующими спецификациями:

1. $U_1 = \{x_1, \dots, x_n\}$ - входной слой, $U_2 = \{R_1, \dots, R_k\}$ - слой нейронов правил, $U_3 = \{c_1, \dots, c_m\}$ - выходной слой классифицирующих нейронов.
2. Каждая связь между узлами $x_i \in U_1$ и $R_r \in U_2$ помечена лингвистическим термом $A_{jr}^{(i)}$ ($j_r \in \{1, \dots, q_i\}$).
3. $W(R, c) \in \{0, 1\}$ определяется для всех $R \in U_2, C \in U_3$.
4. Соединения, которые происходят из одного и того же входного узла x и имеют идентичные метки, одинаковый вес в любой момент времени. Такие соединения называются связанными соединениями, а их вес называется взвешенным.
5. Обозначим $L_{x,R}$ метку соединения между узлами $x \in U_1$ и $R \in U_2$, тогда для всех $R, R' \in U_2$ выполняется

$$(\forall (x \in U_1) L_{x,R} = L_{x,R'}) \Rightarrow R = R'.$$

6. Для всех узлов правил $R \in U_2$ и всех узлов $C, C' \in U_3$ имеем

$$(W(R, C) = 1) \wedge (W(R, C') = 0) \Rightarrow C = C'.$$

7. Для всех выходных узлов $C \in U_3$ выполняется

$$O_c = a_c = NET_c.$$

8. Для всех выходных узлов $C \in U_3$ NET_c рассчитывается по формуле

$$NET_c = \frac{\sum_{R \in U_2} W(R, C) O_R}{\sum_{R \in U_2} W(R, C)}.$$

Система NEFClass может быть построена по неполному знанию об образцах, а затем откалибрована путем обучения или создана с самого начала посредством обучения. Пользователь должен определить количество начальных нечетких множеств, разбивающих интервал значений, в которых лежит признак, а также наибольшее количество узлов правил, созданных в скрытом слое.

Обучение в системе NEFClass

Для обучения была использована гауссовская функция принадлежности [Зайченко, 2004], поскольку она дает более реальные данные и обеспечивает гибкость настройки нейронной сети.

$$\mu(x) = \exp\left[-\left(\frac{x-c}{\sigma}\right)^2\right]$$

Рассмотрим непосредственно алгоритм обучения. Алгоритм обучения системы NEFClass [Зайченко, 2008] состоит из двух этапов обучения. На первом этапе генерируется или задается база лингвистических правил вида «если - то». На втором этапе обучения производится настройка весов нейронной сети с помощью следующих алгоритмов обучения: градиентного, генетического и алгоритма сопряженных градиентов.

Рассмотрим систему NEFClass с n входными нейронами x_1, \dots, x_n , $k \leq k_{max}$ нейронами правил и m выходными нейронами C_1, \dots, C_m . Также задано обучающее множество образцов $L = \{(p_1, t_1), \dots, (p_s, t_s)\}$, каждый из которых состоит из входного образца $p \in R^n$ и желаемого образца $t \in \{0, 1\}^m$.

Целью первого этапа является создать k нейронов правил в системе NEFClass, данный этап состоит из следующих шагов:

1. Выбираем следующий образец (p, t) из L
2. Для каждого входного нейрона $x_i \in U_1$ находим такую функцию принадлежности μ_j , что:

$$\mu_{j_i}^{(i)}(p_i) = \max_{j \in \{1, \dots, q_i\}} \{\mu_{j_i}^{(i)}(x_i)\}$$

3. Если узлов правил осталось меньше, чем k_{max} и не существует узла правила R с $W(x_1, R) = \mu_{j_1}^{(1)}, \dots, W(x_n, R) = \mu_{j_n}^{(n)}$, то создаем такой узел и соединяем его с выходным узлом C_i , если $t_i = 1$.
4. Если в L остались необработанные образцы и $k \leq k_{max}$, то идем в начало и повторяем снова все шаги, иначе останавливаемся.

Целью второго этапа обучения является настройка весов нейронной сети. В данном исследовании были использованы три алгоритма обучения нейронных сетей: градиентный, генетический и алгоритм сопряженных градиентов. Результаты исследования и сравнительная характеристика работы этих алгоритмов приводятся ниже в данной статье.

Вычисление признаков объектов

В данной части статьи рассматриваются признаки, которые станут основой для распознавания зданий на спутниковых изображениях.

В работе рассматривается 3 вида крыш: плоская, круглая и треугольная крыша. Каждый из этих видов крыши представляет различные населенные пункты, отличающиеся местоположением и ландшафтом. Так здания с плоскими крышами преобладают в городских территориях, и соответствуют высокоэтажному строению. Треугольные крыши наиболее часто встречаются в домах загородного типа, распознать на изображении большую часть таких крыш, можно говорить о том, что данный населенный пункт является небольшим и в нем преобладают невысокие строения. Круглые крыши зданий преобладают в

населенных пунктах находящихся в засушливых, пустынных регионах. Кроме того такой вид крыши часто обозначает религиозные сооружения, если вокруг преобладают плоские и/или треугольные крыши. На Рисунке 2 приведены типичные изображения населенных пунктов, которые можно охарактеризовать по форме крыш как город, пригород и населенный пункт в пустынном регионе.



Рис. 2 Изображения населенных пунктов:

(а) - город в пустынном регионе; (б) - пригород; (в), (г) - город

На основе знаний о геометрической форме крыши зданий были подобраны оптимальные признаки для распознавания. Рассмотрим подробнее используемые в работе признаки.

1. Округлость здания для каждого сегмента рассчитывается как отношение площади к периметру. Этот признак может принимать значения от 0 до 1. Вычисляется по следующей формуле:

$$\text{Округлость} = \frac{4\pi * \text{Площадь}}{\text{Периметр}^2}$$

где Площадь сегмента рассчитывается как количество принадлежащих ему пикселей, а Периметр сегмента рассчитывается как количество пикселей лежащих на границе области.

2. Среднее значение углов [Müller, Zaum, 2005]. Углы являются важным признаком, позволяющим дифференцировать здания от других объектов. Для того чтобы определить значение углов здания в выделенном сегменте необходимо построить шестиугольник по следующей схеме: 1. Проводится главная ось между двумя точками контура, которые находятся на максимальном расстоянии друг от друга. 2. Определяются две вспомогательные оси с каждой стороны главной оси. Для этого строим вертикальные линии, лежащие на главной оси и имеющие максимальное расстояние от контура до главной оси. 3. Проводим поперечную ось, это должна быть вертикальная линия на главной оси, которая соединяет две точки контура находящиеся на максимальном расстоянии друг от друга. 4. Полученные шесть точек соединяем линиями в шестиугольник. В качестве признаков возьмем среднее значение двух углов, находящиеся по обе стороны главной оси.
3. Длина здания определяется как отношение длины главной оси к длине поперечной оси полученного выше шестиугольника [Müller, Zaum, 2005]:

$$\text{Длина} = \frac{m_{20} + m_{02} + \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2}}{m_{20} + m_{02} - \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2}}$$

4. Полнота вычисляется как отношение общего количества пикселей, которые лежат внутри шестиугольника и не принадлежат к исследуемой области к количеству пикселей, которые лежат за пределами шестиугольника и принадлежат исследуемому сегменту.
5. Градиент яркости поверхности крыши. Для определения градиента яркости будет использован оператор Собеля, который вычисляет приближенное значение градиента яркости изображения. В результате применения оператора Собеля в каждой точке изображения получим вектор градиента яркости в этой точке. Таким образом, будет найдено направление наибольшего увеличения яркости и величина её изменения в этом направлении. В разработанном алгоритме изображение делится на 20 сегментов, значение оператора Собеля вычисляется в центре каждого такого сегмента. В зависимости от полученных направлений увеличения яркости можно говорить об одном из видов крыш.

Экспериментальные исследования

Для проверки эффективности работы рассматриваемой нейронной сети были проведены тесты на изображениях с различными видами крыш. Для проведения экспериментов было использовано 90 изображений, из них 60 изображений использовались для обучения нейронной сети. Изображения с круглыми крышами брались из регионов с преобладанием пустынного ландшафта, треугольные - из пригородных территорий, а изображения с плоскими крышами – из территорий крупных городов. Для тестирования было использовано 30 изображений.

В разработанной системе NEFClass входам нейронной сети соответствуют описанные выше 5 признаков: округлость, среднее двух углов, лежащих по обе стороны главной оси, построенного шестиугольника, длина здания, полнота, градиент яркости поверхности крыши. Выходам сети соответствуют 3 вида крыш: круглая, треугольная, плоская.

Эксперимент был поставлен следующим образом. Каждое изображение было предварительно обработано и сегментировано на отдельные образцы – паттерны. Каждый такой паттерн содержит изображение крыши одного из видов зданий. Для генерации базы правил каждому паттерну на изображении ставится в соответствие подходящая форма крыши. Таким образом, создается база правил. Проводим обучение нейронной сети одним из выбранных методов: генетическим, градиентным или методом сопряженных градиентов. Далее на обученную сеть подаем тестовое изображение, на выходе получаем изображение, на котором отмечены распознанные крыши различными цветами, в зависимости от формы крыши. Так круглые крыши отмечены желтым цветом, треугольные – синим, а плоские – красным.

Если сравнить реализованные алгоритмы обучения между собой, то можно увидеть, что не все три метода справляются хорошо и можно отметить следующие тенденции. Метод сопряженных градиентов в начале обучения сходится значительно быстрее, чем обычный градиентный метод и в итоге дает лучший результат. Генетический же метод сходится гораздо медленнее градиентного метода и метода сопряженных градиентов. Кроме того, генетический метод отличается этапным улучшением суммарной погрешности, то есть при общей тенденции сходимости и медленном уменьшении ошибки, есть моменты довольно резкого уменьшения ошибки. Необходимо отметить, что градиентный метод и метод сопряженных градиентов сошлись к ошибке, которая была ниже установленного порога, поэтому оказались лучше генетического алгоритма. Однако в случае большего количества итераций и, возможно, других данных генетический алгоритм мог оказаться лучше, о чем свидетельствуют результаты, полученные при выполнении работы и теоретические данные [Зайченко, 2004].

Сказанное выше проиллюстрировано на Рис. 3, который показывает зависимость суммарной погрешности от пройденных итераций по каждому из выбранных методов.

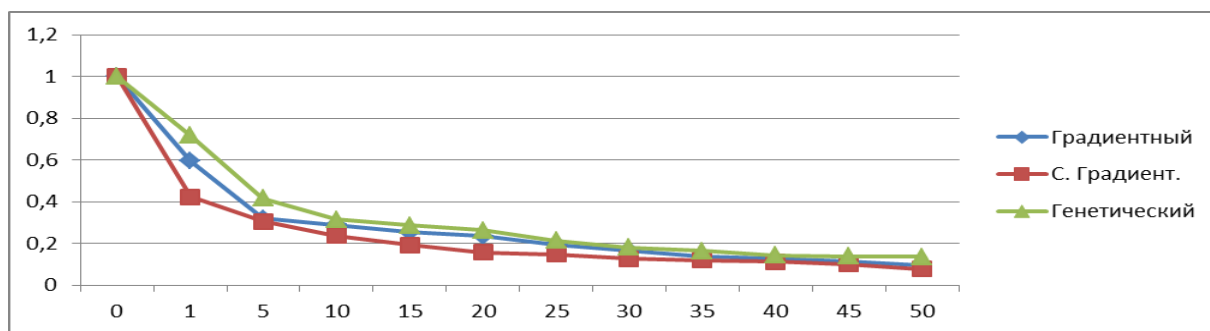


Рис.3. Графическое представление зависимости суммарной погрешности от пройденных итераций

Рассмотрим график на Рис. 4, который показывает процесс обучения нейронной сети для каждого вида крыши, на примере обучения методом сопряженных градиентов, поскольку этот метод дал наилучший результат.

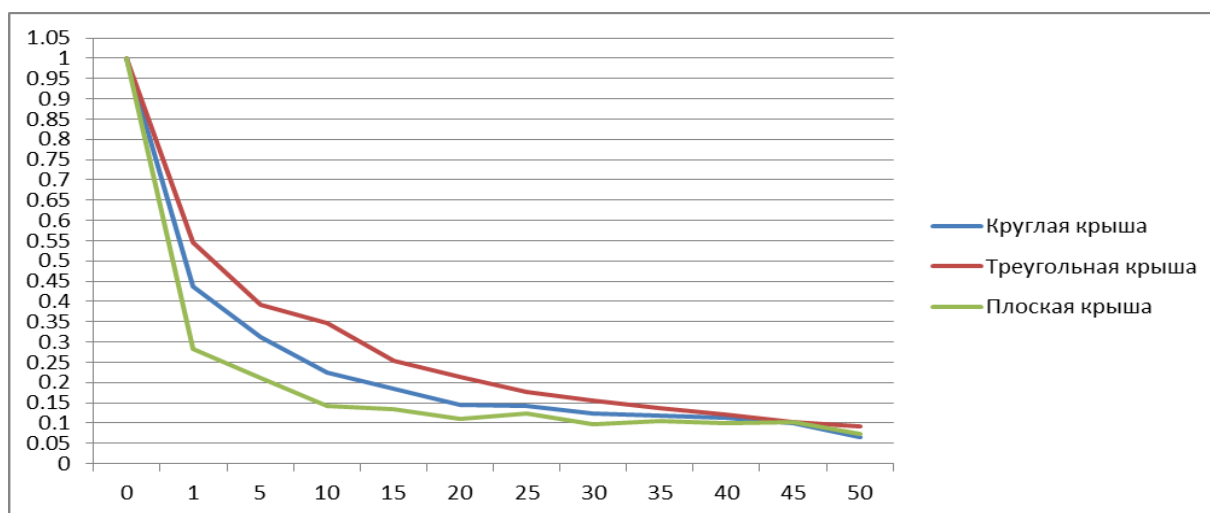


Рис.4. Графическое представление зависимости погрешности от пройденных итераций для каждого вида крыши

Можно увидеть, что наиболее легким примером для обучения нейронной сети были здания с плоскими крышами. Среди зданий этого типа наибольшее количество распознанных. В тоже время треугольная форма крыши была наиболее сложным примером для обучения и дольше других удерживала значение погрешности выше желаемого порога. Также треугольная форма крыши чаще всего встречалась в нераспознанных образцах, что можно объяснить более сложной геометрической формой.

Для эксперимента было отобрано 90 изображений – 60 для обучения и 30 для тестирования. Результаты обучения и классификации приведены в Табл. 1.

Алгоритмы обучения весовых коэффициентов	Обучение		Тестирование	
	Величина ошибки	% ошибочно классифицированных	Величина ошибки	% ошибочно классифицированных
Генетический	0,13	0	0,19	23,3
Градиентный	0,09	0	0,12	13,3
Сопряженного градиента	0,07	0	0,09	6,6

Табл. 1. Результаты обучения и тестирования

По полученным результатам эксперимента можно сделать вывод, что наилучший результат был получен при обучении с помощью метода сопряженных градиентов. Необходимо отметить, что система не всегда правильно определяет тип крыши, что говорит о необходимости коррекции используемых признаков. Кроме того, наличие вокруг дома на близком расстоянии в большом количестве деревьев увеличивает процент ошибочно классифицированных образцов. Это верно как для критерия количества неправильно классифицированных символов, так и по скорости обучения. При тестировании было ошибочно классифицировано 14,4 % из проверочного набора данных.

Выводы и перспективы дальнейших исследований

Практическая реализация модели NEFClass показала, что эта система дает достаточно высокие результаты применительно к задаче распознавания зданий на спутниковых снимках. Программное обеспечение, разработанное при выполнении работы, показывает основные возможности нечетких нейронных сетей типа NEFClass и может служить базой для развития темы распознавания зданий в дальнейшем. Кроме того, разработанное программное обеспечение может быть использовано для проведения экспериментов по совершенствованию построения и обучения нечетких нейронных сетей. Реализовав на практике градиентный, генетический алгоритмы, и алгоритм сопряженных градиентов, и проанализировав полученные данные, можно заключить, что лучшим для решения поставленной перед системой задачи оказался метод сопряженных градиентов, который быстро сходится и дает наилучшие результаты. Необходимо отметить, что требуются дальнейшие исследования, направленные на усовершенствования нейросетевой архитектуры и методов ее обучения, об этом говорит достаточно высокий процент ошибочно классифицированных зданий. Это улучшение может быть получено путём коррекции признаков, которые поступают на вход сети и увеличения обучающей выборки. Также возможно улучшение результата путем дообучения базы правил с помощью алгоритма «наилучший для каждого класса» [Зайченко, 2008]. Такой подход позволит снизить процент неверно классифицированных зданий. В заключение следует ещё раз отметить, что по результатам проведенных исследований можно говорить о достаточно высокой результативности применения нечеткого классификатора к задачам распознавания зданий.

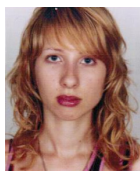
Благодарности

Статья частично финансирована из проекта ITNEA XXI Института Информационных теорий и Приложений FOI ITNEA и консорциума FOI Bulgaria (www.ithea.org, www.foibg.com)

Литература

1. Зайченко Ю.П. Основы проектирования интеллектуальных систем. Учебное пособие. – К.: «Издательский Дом «Слово»», 2004. – С. 352
2. Зайченко Ю.П. Нечёткие модели и методы в интеллектуальных системах. Учебное пособие для студентов высших учебных заведений. – К.: «Издательский Дом «Слово»», 2008. – С. 344
3. S. M'uller and D. W. Zaum, 2005, Robust building detection from aerial images, IAPRS, Vol. XXXVI, Part 3/W24, Vienna, Austria.

Информация про автора



Дьяконова Светлана Владимировна – аспирантка Национального технического университета Украины «КПИ», адрес электронной почты: edeta@bigmir.net

Основные сферы научных исследований автора: применение нечеткого классификатора к задачам распознавания зданий на спутниковых снимках.

ВЕРИФИКАЦИЯ ОПТИКО-ПЕРЕМЕННОГО ОБЪЕКТА ПУТЁМ ОБНАРУЖЕНИЯ ИЗМЕНЕНИЯ ЦВЕТА

Екатерина Горшкова, Роман Телятников, Иван Шумский

Аннотация: Рассматривается задача обнаружения изменения цвета на оптико-переменных объектах. Для решения используется цветовая модель HSL, выработан критерий оценки различия цветов.

Ключевые слова: цветовая модель, цветовой тон, изменение цвета.

Введение

Одним из способов автоматической проверки подлинности документов, удостоверяющих личность, является верификация оптико-переменного объекта по его фотографическим изображениям. Под объектом понимается изображение, выполненное специальными красками, меняющими цвет в зависимости от угла зрения или от угла освещения. Если изменения цвета объекта не наблюдается, то можно сделать вывод о том, что документ поддельный. Для человеческого глаза это не составляет большого труда. Человек смотрит на изображение под одним углом зрения и видит, например, красный цвет, далее, меняя угол зрения, видит, например, зелёный цвет (Рис.1), и, зная, что на данном типе документов объект имеет именно такие цвета, он делает вывод, что документ подлинный.



Рис.1. Изображение паспорта с оптико-переменным объектом

Чтобы заставить машину решить такую тривиальную для человека задачу, требуется ответить на ряд вопросов:

- в какой цветовой модели работать?
- что понимать под цветными пикселями?
- каким может быть разброс цвета?
- какие цвета считать разными?

Выбор цветовой модели

Нужно выбрать такую цветовую модель, которая лучше других подходит для детекции изменения цвета оптико-переменного объекта. Ещё одним условием является устойчивость к изменению яркости изображения, так как для получения изображения могут использоваться разные камеры и различные условия освещения.

В зависимости от поставленной задачи могут использоваться аддитивные (RGB), субтрактивные (CMY) и перцепционные (HSL, HSV) цветовые модели. Каждая из них хороша в определенной области задач, поэтому однозначного предпочтения какой-то из них отдать нельзя.

RGB и CMY - довольно распространенные модели. Первая часто используемая в приборах, излучающих свет, таких, например, как мониторы, проекторы и другие подобные устройства, вторая - для печати [Корабельникова, 2008]. Для данной задачи они не подходят, потому как такие понятия как цветовой тон и насыщенность для данных моделей являются расчётными величинами, и определяются значениями всех трёх компонент. Ещё один недостаток – отсутствие отдельной составляющей по яркости.

HSL и HSV хорошо приспособлены для описания цветов таким образом, как это свойственно человеку. Глядя на окрашенный объект, человеку проще его описать с помощью цвета, яркости и насыщенности, что и делают данные цветовые модели [Гонсалес, 2006]. Несомненным преимуществом данных моделей при построении алгоритмов обработки изображений является простота понимания, поскольку в их основе лежит естественное и интуитивно близкое человеку описание цвета, а ведь именно человек, в конечном счёте, является и разработчиком и пользователем этих алгоритмов. Цветовой тон, который характеризует собственно цвет (чистый желтый, оранжевый, красный и т.д.) обе модели описывают одинаково, поэтому в данном случае не важно, какую модель использовать. К достоинствам этих моделей можно также отнести и то, что за цветовой тон отвечает лишь одна компонента, что значительно упрощает вычисления, а также довольно простой переход из пространства RGB. При переводе значений из RGB в HSV или HSL общее количество возможных цветов уменьшается, и их становится из 16 млн. всего 3 млн. Однако применительно к нашей задаче это не является недостатком, ведь когда человек смотрит на изображение опико-переменного объекта, ему надо выделить основной цветовой тон, не вдаваясь в детали, поэтому он обобщает цвет объекта.

Lab - равноконтрастная модель с точки зрения человеческого восприятия. Однако человеческое восприятие, и то, что видит камера - это не одно и то же. Системы HSL, HSV получаются прямым преобразованием из RGB, т.е. отражают то, что видит камера. В то время как при получении Lab производится пересчёт сначала в XYZ, а потом уже в Lab, и тем самым вносятся искажения в исходные данные. К тому же формулы перевода в Lab верны при использовании стандартизованного CIE D65 дневного источника освещения [Fairchild, 2005], а значит подсветка, используемая при получении изображений так же будет вносить искажения.

Исходя из описанных выше рассуждений, была выбрана именно модель HSL.

Что считать “цветным”?

Наиболее натренированные наблюдатели способны различать по цветовым тонам 160 цветов, по насыщенности — 25, по светлоте — не более 64 [Прохоров, 1988].

Если взять два совершенно разных цветовых оттенка с низкой насыщенностью, то человеческий глаз увидит серые цвета, и определить цветовой тон уже не сможет (Рис.2). То же самое происходит при низкой светлоте (Рис.3), а при высокой светлоте получаются бледные оттенки близкие к белому, для которых определить цветовой тон также не представляется возможным для человека (Рис.4). А задача состоит в том, чтобы обнаружить изменение цвета, которое видно человеческому глазу.

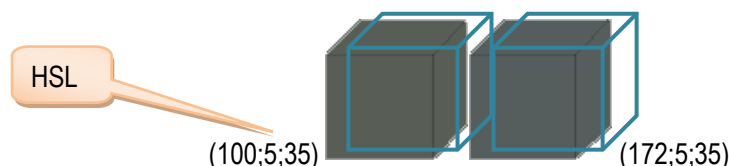


Рис.2. Изображение объектов в системе HSL с низкой насыщенностью, но разным цветовым тоном

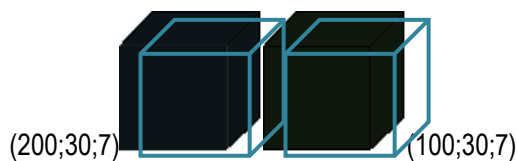


Рис.3. Изображение объектов в системе HSL с низкой светлотой, но разным цветовым тоном

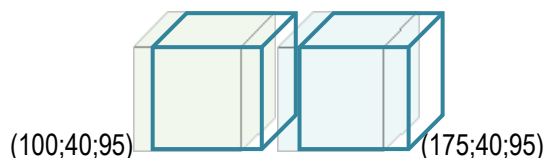


Рис.4. Изображение объектов в системе HSL с высокой светлотой, но разным цветовым тоном

Поэтому после перехода в цветовую модель HSL нужно оставить лишь те пиксели, изменение цвета которых уловимо для человеческого глаза. *Опытным путём было получено, что если насыщенность < 10%, а светлота < 15% или > 85%, то такие пиксели нужно отбросить из рассмотрения, следовательно, оставшиеся пиксели будут считаться цветными.*

Определение цветового тона

На следующем шаге нужно определиться с разбросом цветового тона. Для этого необходимо построить его гистограмму. На рисунке 5 видно, что разброс одного цветового тона отличается от разброса другого цветового тона.

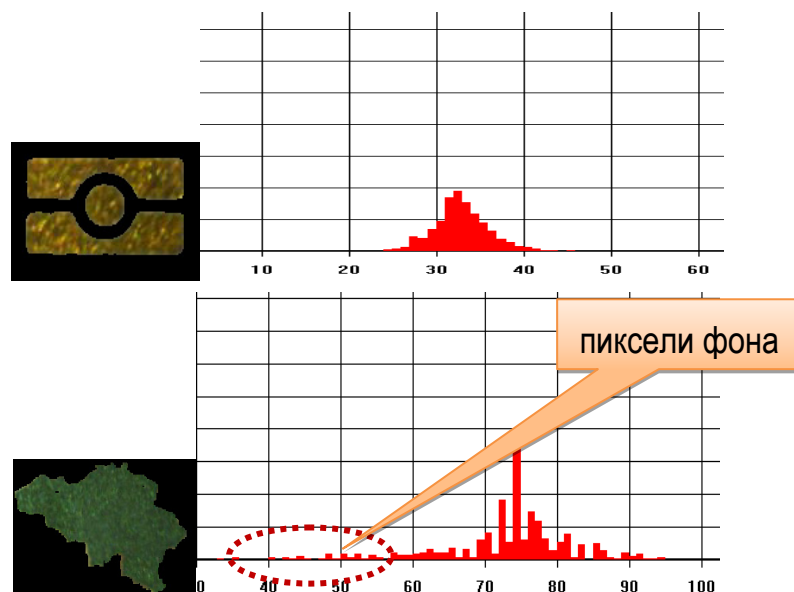


Рис.5. Гистограммы цветового тона разных оптико-переменных элементов

К сожалению, при выделении мелких деталей, в результирующем изображении могут присутствовать и пиксели фона, что конечно влияет на величину цветового разброса.

Был проведён анализ оптических свойств оптико-переменных объектов на большом числе образцов документов. По результатам оценки гистограмм цветового тона было установлено, что диапазон значений по данной цветовой составляющей для одного элемента может составлять от 14 до 56 градусов. На ширину данного диапазона помимо оптических свойств элемента влияет цвет фона, полностью избавиться от которого на этапе выделения объекта не всегда представляется возможным. Поэтому было

введено понятие цветового окна - это диапазон минимальной ширины, в который попадает не менее 75% от общего числа всех цветных пикселей. Человек видит цвет элемента и оценивает его усреднённое значение, а машина вычисляет цветовое окно. В случае, если 75 % пикселей распределено в интервале большем, чем цветовое окно, считается, что элемент не имеет равномерного цвета, что является отклонением от стандарта. Поэтому данные элементы считаются поддельными.

Критерий различия цветового тона

Теперь необходимо определиться, что считать разными цветами. На рисунке 6 отчётливо видно, что гистограммы оптико-переменного элемента, отсканированного под разными углами освещения, не перекрываются, что говорит о возможности их уверенного различия по цветовому тону.

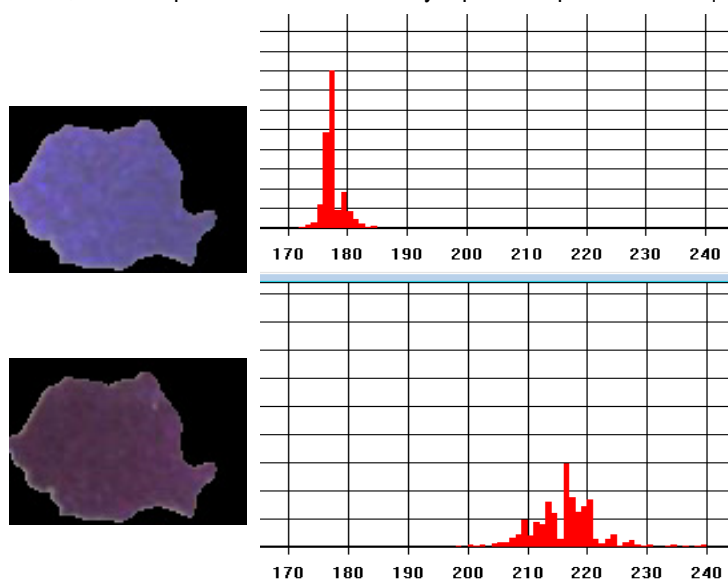


Рис.6. Гистограммы цветового тона одного оптико-переменного элемента под разным углом освещения

Также могут быть варианты, где гистограммы перекрываются на 5-10 процентов от площади всех цветных пикселей (Рис.7). По результатам оценки степени пересечения гистограмм цветовых тонов было принято решение считать цвета различными в случае, если площадь пересечения составляет не более 15% от количества всех цветных пикселей.

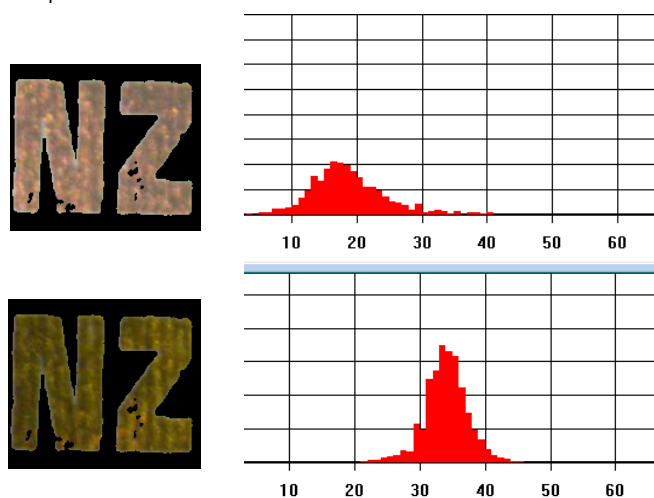


Рис.7. Перекрывающиеся гистограммы цветового тона одного оптико-переменного элемента под разным углом освещения

Заключение

В работе была рассмотрена задача автоматической проверки подлинности документа, путём обнаружения изменения цвета на оптико-переменных объектах. Был предложен метод решения данной задачи, основанный на анализе и сравнении гистограмм цветового тона с использованием модели HSL.

Также были рассмотрены вопросы определения допустимых параметров цветных пикселей, таких как величина разброса цветового тона, минимально допустимые значения яркости и насыщенности. Данный метод показал свою эффективность и был внедрен в существующую систему проверки подлинности документов.

Библиография

[Корабельникова, 2008] Г. Корабельникова. Цвет и цветовые модели. <http://www.woweb.ru/publ/22-1-0-144>, 2008

[Гонсалес, 2006] Гонсалес Р., Вудс Р. Цифровая Обработка Изображений. Техносфера.Москва, 2006 - 432с.

[Fairchild, 2005]Mark D. Fairchild. Color Appearance Models, 2nd Edition. John Wiley & Sons, 2005. – 408 p.

[Прохоров, 1988] А.М. Прохоров. Физическая энциклопедия. Москва, Советская энциклопедия, 1988, т.5. - 420 с.

Сведения об авторах



Горшкова Екатерина – инженер-программист, ООО «Регула», 220036 ул. Волоха 1-314, г. Минск, Беларусь, e-mail: katsiaryna.harshkova@regula.by

Основные направления деятельности: обработка цветных изображений, элементы защиты и системы проверки подлинности документов



Телятников Роман – к.т.н., научный руководитель разработками ПО, ООО «Регула», 220036 ул. Волоха 1-314, г. Минск, Беларусь, e-mail: raman.tsialatnikau@regula.by

Основные направления деятельности: распознавание образов, обработка изображений, нейрофизиология



Шумский Иван – к.т.н., генеральный директор ООО «Регула», 220036 ул. Волоха 1-314, г. Минск, Беларусь, e-mail: ivan.shumsky@regula.by

Основные направления деятельности: автоматизация анализа подлинности документов и денежных банкнот: проектирование оборудования и программного обеспечения

ВЕРИФИКАЦИЯ ЦВЕТНЫХ ОБЪЕКТОВ НА СЛОЖНЫХ СЦЕНАХ

Анатолий Протосавицкий, Екатерина Горшкова,
Дмитрий Пилютник, Роман Телятников, Иван Шумский

Аннотация: В данном докладе представлен метод описания цветных объектов на сложных сценах, а также алгоритмы сравнения таких объектов в различных условиях освещения. Описанные алгоритмы были разработаны после анализа большого количества различных изображений в нескольких цветовых моделях при решении задачи распознавания объектов. Размеры объектов, их форма и цвет могут быть произвольными. Методы выделения объектов могут отличаться от рассматриваемых.

Ключевые слова: цветные объекты, освещённость.

ACM Classification Keywords: algorithms, verification

Введение

На данный момент всё большее распространение получают цветные цифровые фото- и видеокамеры. Это и системы видео наблюдения, и медицинское оборудование, и многое другое. Их использование в свою очередь приводит к необходимости анализа объектов на сложных цветных сценах. Эффективные методы описания цветных объектов необходимы при решении таких задач как распознавание, верификация и многих других. Существуют различные подходы и методы для описания и выделения объектов произвольной формы, размеров и цвета, и под каждую задачу разрабатываются алгоритмы, которые дают приемлемые результаты для конкретных условий их эксплуатации. Причинами, которые могут влиять на устойчивость работы таких систем, являются изменения условий освещённости (его интенсивность и спектр), цветовой разброс однотипных объектов и пр.

Представленный в данной работе алгоритм позволяет производить сравнение объектов при различной освещённости, что значительно повышает устойчивость работы подобных систем. Также он позволяет настраивать чувствительность к разбросу цветового тона, яркости и насыщенности. Разработка алгоритма производилась для решения задачи распознавания и верификации документов (паспортов, идентификационных карт и др.) по их цветным фотографическим изображениям. Особенностями данной задачи является то, что местоположение цветных объектов на изображении является постоянным, что позволяет использовать простую схему описания изображения документа в целом.

На изображениях документов зачастую присутствуют блики, голограммы и другие элементы, которые делают невозможным выделение структурных особенностей. К тому же многие документы имеют схожий формат заполнения, и отличаются лишь по цвету. Указанные факторы ухудшают работу структурных алгоритмов распознавания. Именно поэтому возникла задача сравнения цветных изображений, вариант решения которой представлен в данном докладе.

Обоснование выбора цветовой модели

Не секрет, что большинство современных цветовых моделей ориентированы либо на устройства цветовоспроизведения, либо на определённые прикладные задачи, когда работа с цветом является непосредственной целью. [Гонсалес Вудс, 2006] Аппаратно-ориентированными цветовыми моделями, наиболее часто используемыми на практике, являются модель RGB для цветных мониторов и широкого класса цветных видеокамер, модели CMY и CMYK для цветных принтеров. Для нашей задачи более подходящими являются другие цветовые модели, такие как HLS и HSV. Одним из преимуществ этих

моделей является лучшее соответствие восприятию цвета человеком. Например, описывая цвет автомобиля, человек не говорит о процентном содержании в нём красного(R), зелёного(G) и синего(B). Человек описывает объект с помощью цвета (цветового тона), насыщенности и светлоты либо яркости. Вторым неоспоримым преимуществом является разделение яркостной и цветовой составляющей на изображении, что даёт возможность использовать многие из полутоновых методов обработки для решения тех или иных задач. Особенности данных моделей таковы, что с каналами можно работать независимо друг от друга. Коническое представление выбранной нами модели HSVпредставлено на рисунке 1.

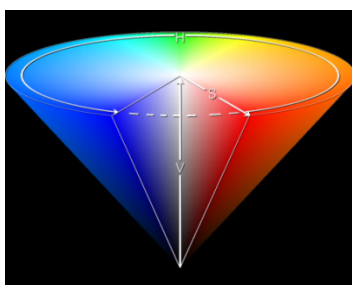


Рис 1.

Недостатком данных цветковых моделей является ошибка вычисления насыщенности при переводе из RGB. Для системы HLSэто проявляется как при значениях RGB близких к 0, так и близких к 255, а для HSV– лишь для близких к 0 значений. Это означает, что степень доверия насыщенности при таких условиях не должна быть высокой.

Для определения различия цветов можно использовать так же цветовую модель Lab. В данной модели так же присутствует отдельная координата яркости. Значения цветового тона, рассчитываемые в системе Labна основании координат а и b, представлены на рисунке 2.

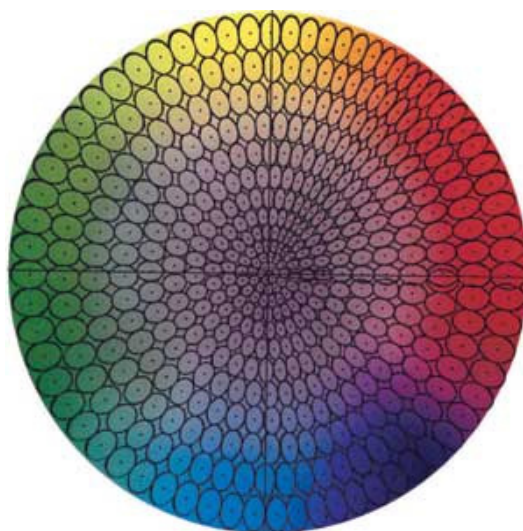


Рис. 2

Как известно, при разработке Lab преследовалась цель создания цветового пространства, изменения цвета в котором будет более линейным с точки зрения человеческого восприятия, то есть с тем, чтобы одинаковое изменение значений координат в разных областях цветового пространства производило одинаковое ощущение изменения цвета.[ГонсалесВудс, 2006] Однако стоит отметить, что отличие от HSV(HLS)формулы перевода в систему Labне являются линейными, что безусловно вносит некоторые искажения. К тому же все расстояния между цветами в Labрассчитаны для идеального источника света CIE D65.[Fairchild, 2005]Одной из целей проведённой работы являлось создание такого алгоритма сравнения, который был бы нечувствителен к изменению освещённости. Результаты оценки на изображениях с разным уровнем освещённости показали, что и в системе HSV(HLS), и в системе Lab цветовой тон практически неизменен. Графики изменения значения цветового тона для шести различных цветов при постепенно возрастающем уровне освещённости показаны на рисунке 3.

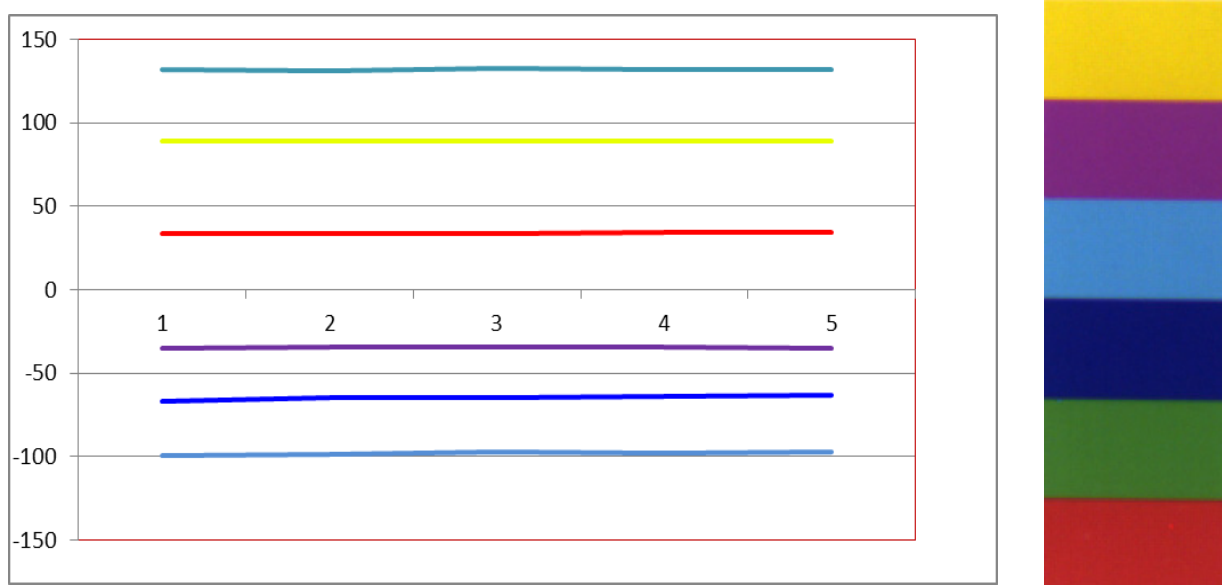


Рис. 3

Таким образом, можно сделать вывод о том, что свойство независимости цветового тона к уровню освещённости соблюдается для каждой из представленных цветовых моделей. Т.е. для сравнения цветов можно использовать любую из них. Отличия в результате будут обусловлены тем, что Lab разработана как равноконтрастная модель, и результаты её работы должны быть более понятны для человеческого восприятия. Однако линейность формул перевода и простота пониманияпредопределили выбор в пользу модели HSV.

Описание объекта (сцены)

Стоит отметить, что способ описания в большей степени зависит от решаемой задачи. Важное значение имеют такие факторы как форма и размеры объектов, их контраст относительно сцены, постоянство различных параметров объекта и сцены. Под каждую конкретную задачу можно построить такое описание, которое будет наиболее полно соответствовать существенным особенностям объекта (сцены). Для объяснения работы метода будет использоваться довольно простое описание, которое позволяет наглядно продемонстрировать работу предлагаемого алгоритма.

Для получения информации о цвете требуются значения по каналам цветовой модели, а также параметры их изменения, которые и будут описывать объект. Построение описания цветного объекта (сцены) будет проводиться согласно схеме, представленной на рисунке 4.

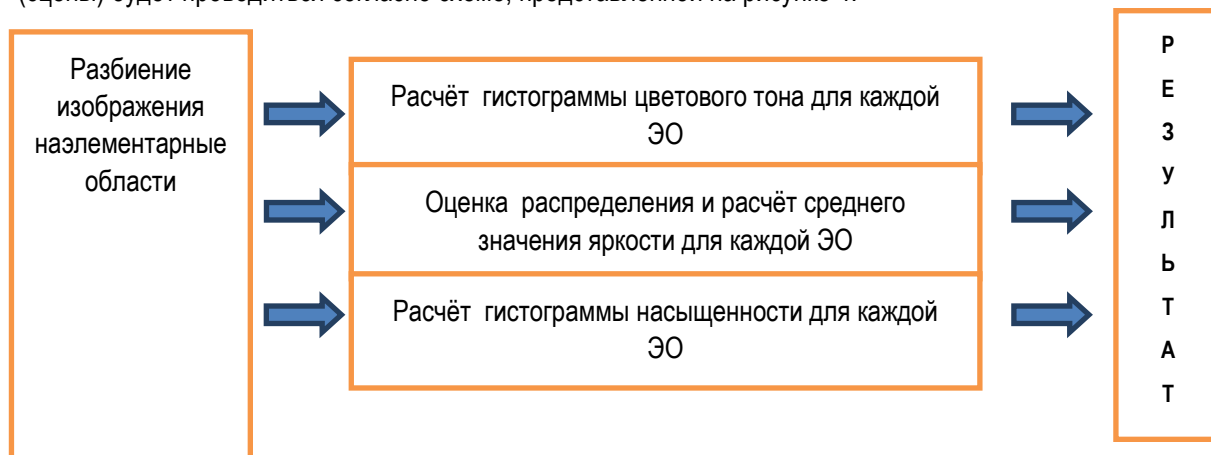


Рис. 4

Первым этапом получения описания является разбиение изображения на некоторое количество одинаковых по размеру прямоугольных областей (далее – элементарная область или ЭО). Другими словами на изображение накладывается прямоугольная сетка, разбивающая его на мелкие области.

Далее, рассчитываются гистограммы цветового тона, яркости и насыщенности для каждой ЭО. Расчёт **гистограммы цветового тона** имеет ряд особенностей. Весь диапазон разбивается на некоторое количество равных интервалов. Это позволяет несколько уменьшить влияние несущественных различий между одинаковыми объектами на разных изображениях. В зависимости от задачи данным параметром можно влиять на точность сравнения, увеличивая или уменьшая количество таких интервалов. В каждое значение гистограммы записывается количество пикселей, попадающих в соответствующий цветовой диапазон. Общий пример гистограммы цветового тона представлен на рисунке 5.

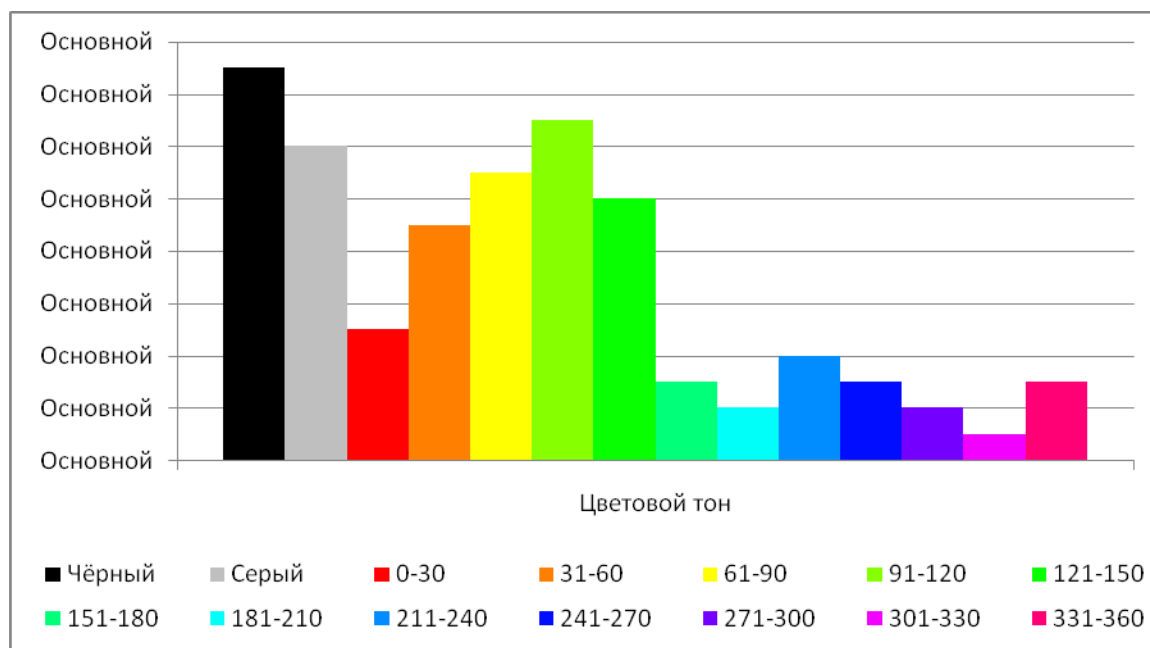


Рис. 5

Помимо этого в гистограмму вводятся два дополнительных столбца. Первый из них – это количество «чёрных» пикселей, т.е. пикселей с низкой яркостью. Второй дополнительный столбец гистограммы предназначен для «серых» пикселей – для тех, у которых значения насыщенности являются недостаточно высокими. Для каждой задачи задаются значения яркости и насыщенности, начиная с которых разработчик считает целесообразным оценивать цветовую составляющую, т.к. при низких значениях яркости или насыщенности даже значительное изменение цветового тона для человека будет неразличимо. [Прохоров, 1988]

Формирование столбца «серых» пикселей происходит по следующим правилам:

пиксели, имеющие значения насыщенности меньше нижней границы интервала, относятся к «серым»

пиксели, имеющие значения насыщенности больше верхней границы интервала, относятся к соответствующему им цветовому интервалу

пиксели, насыщенность которых попадает внутрь заданного интервала, распределяются между столбцом «серых» и «цветовых» пикселей с коэффициентами, определяемыми формулами (1) и (2):

$$K_{color} = \frac{S_{cur} - S_{min}}{S_{max} - S_{min}} \quad (1)$$

$$K_{gray} = 1 - K_{color} \quad (2),$$

где S_{cur} – текущее значение насыщенности, S_{min} – нижняя граница интервала, S_{max} – верхняя граница интервала.

Другими словами, при составлении гистограммы цветового тона считается, что пиксели, имеющие значение насыщенности меньше минимального – это пиксели серого цвета. В решаемой задаче использовались значения $S_{min} = 10$, $S_{max} = 15$.

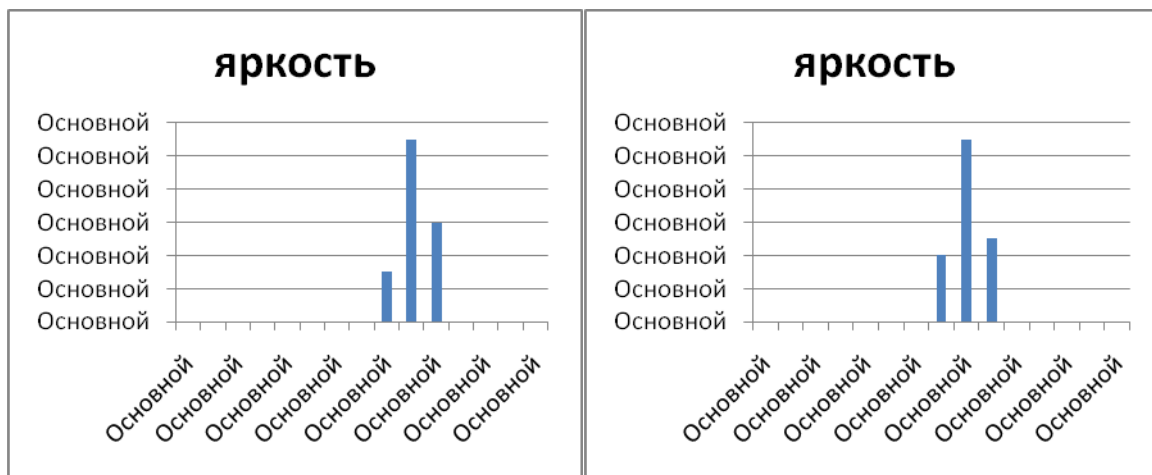


Рис. 6

Аналогичным образом производится расчёт столбца «чёрных» пикселей. Данные интервалы перехода вводятся для исключения ситуации значительного различия гистограмм при значениях яркости близких к граничным. Для примера на рисунке 6 представлены гистограммы канала яркости для двух элементарных областей, имеющих одинаковый цветовой тон, например зелёный.

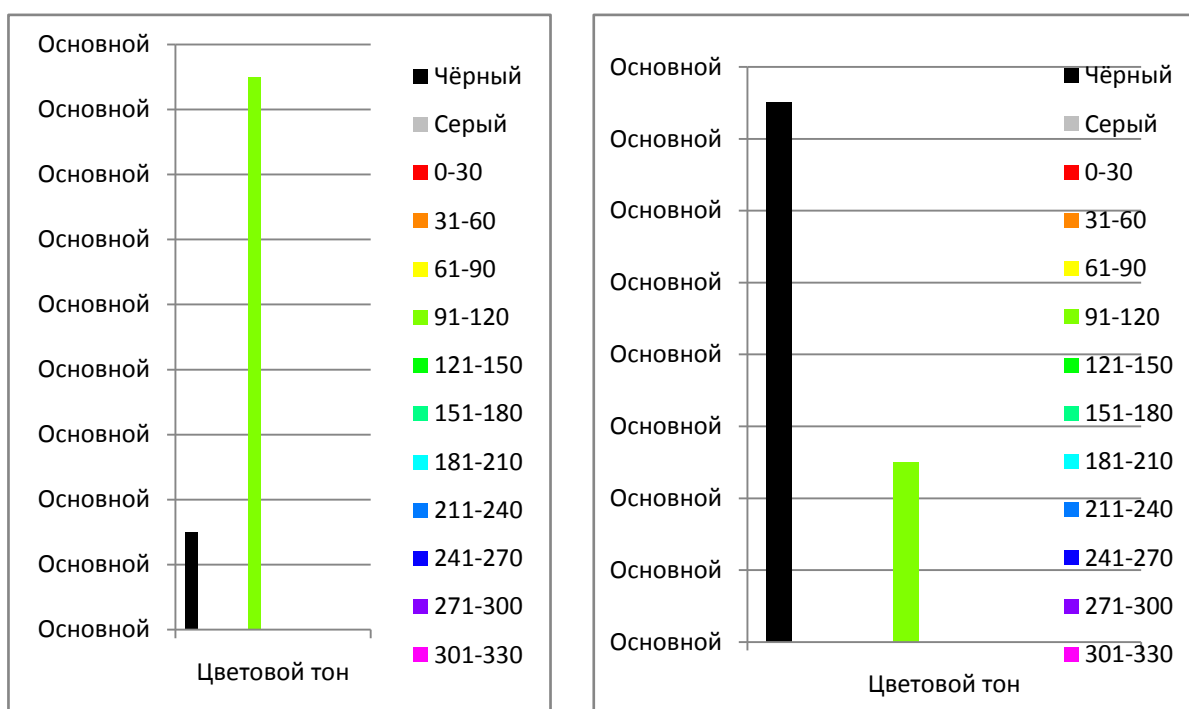


Рис. 7

Если фиксированный порог яркости для данного класса изображений будет равным 10, то для рассматриваемых элементарных областей гистограммы цветового тона будет выглядеть так, как представлено на рисунке 7. Результат пересечения получится равным 40% площади области.

Если же использовать интервал яркости например от 8 до 12, и рассчитать гистограмму с учётом вышеуказанных формул, то процент совпадения составит уже 75%.

Насыщенность для элементарной области описывается гистограммой с настраиваемым количеством отсчётов, что позволяет сократить время расчёта и исключить влияние небольших различий. Особенностью системы HSV является значительное изменение насыщенности при низких значениях яркости. Поэтому при расчёте данной гистограммы так же используется анализ яркости и вводится интервал значений перехода между столбцом «чёрных» пикселей и цветными.

Для каждой ЭО также производится расчёт **гистограммы яркости** и рассчитывается среднее значение. Стоит отметить, что оно считается только для областей, имеющих распределение, близкое к нормальному. Решение о виде распределения принимается на основе анализа критерия согласия Пирсона (критерий Хи-квадрат), для расчёта которого используются гистограмма яркости и эталонное распределение Гаусса того же масштаба. Также исключаются из рассмотрения элементарные области с большим значением динамического диапазона яркости. Данная ситуация является недостатком выбранного способа разбиения изображения. При применении алгоритмов, способных выделять объекты с однородными уровнем яркости, данный недостаток может быть исключён.

Анализ пространственного распределения яркости

Как упоминалось выше, одной из проблем анализа цветных изображений является изменение цвета в зависимости от уровня освещённости. Однако для задач верификации необходимо учитывать эти изменения, ведь по сути объект остаётся тем же, меняются лишь условия его съёмки. То же можно

сказать о системах видеонаблюдения – в случаях, когда солнце заходит за тучи, и когда небо ясное, сцена остаётся той же, меняется лишь уровень освещённости.

Для начала рассмотрим, как меняется значение координаты яркости для приведённых на рисунке 8 однородных серых областей при последовательном изменении уровня освещённости.

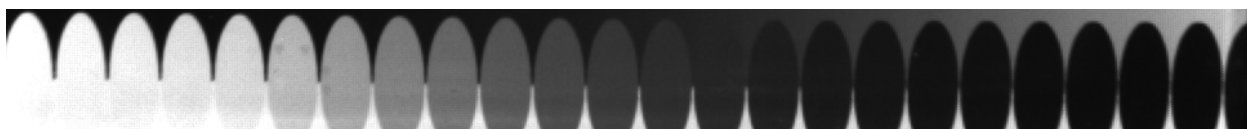


Рис. 8

На рисунке 9 представлены графики изменения параметра Value в модели HSV.

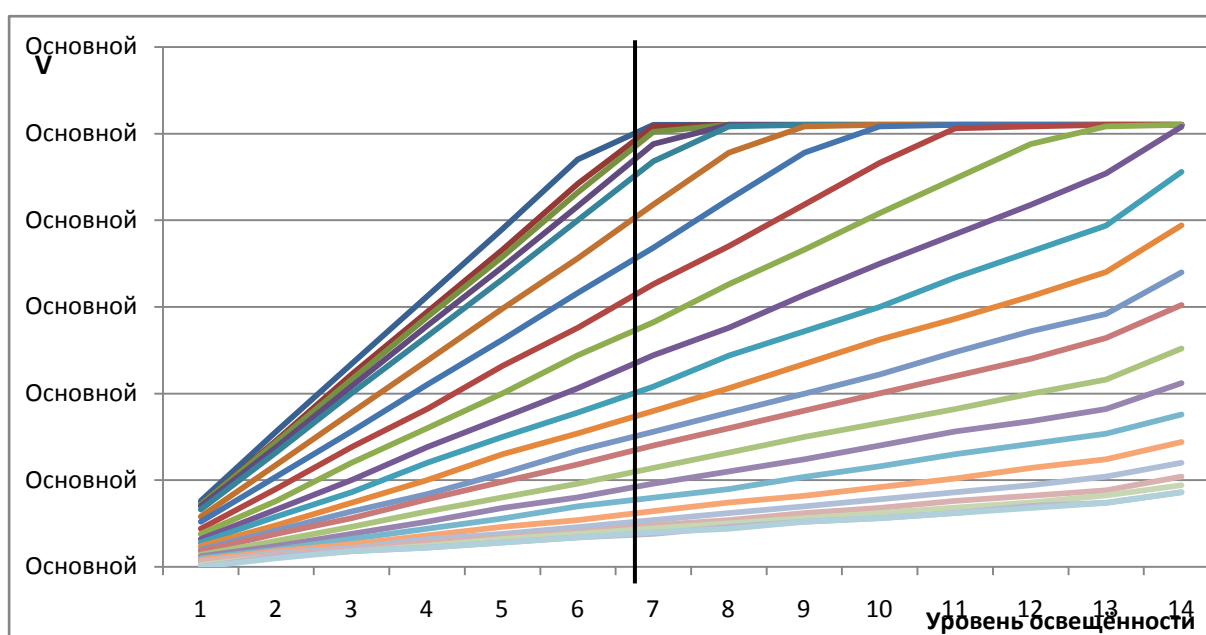


Рис. 9

Из графиков видно, что при увеличении уровня освещённости средняя яркость в области меняется линейно вплоть до достижения уровня насыщения. Для расчётов в качестве отправной точки был выбран такой уровень освещённости, при котором самая яркая область близка к насыщению (уровень 7 на Рис. 9). Примем, что при этом уровне каждая область имеет яркость, которая считается опорной ($L_{оп}$), т.е. яркостью самой области. Тогда закон изменения яркости i -й области при изменении её освещённости описывается следующим выражением:

$$V_i \approx L_{опi} * K(Br) \quad (3),$$

где Br – уровень освещённости, $K(Br)$ – коэффициент изменения яркости при уровне освещённости Br , i – индекс области.

Было установлено, что для всех анализируемых областей значения $K(Br)$ попадают в узкий диапазон значений. Из этого следует, что вне зависимости от $L_{оп}$ при изменении уровня освещённости яркость каждой из областей меняется пропорционально. На рисунке 10 приведены зависимости $K(Br)$ от $L_{оп}$ при различных уровнях освещённости.

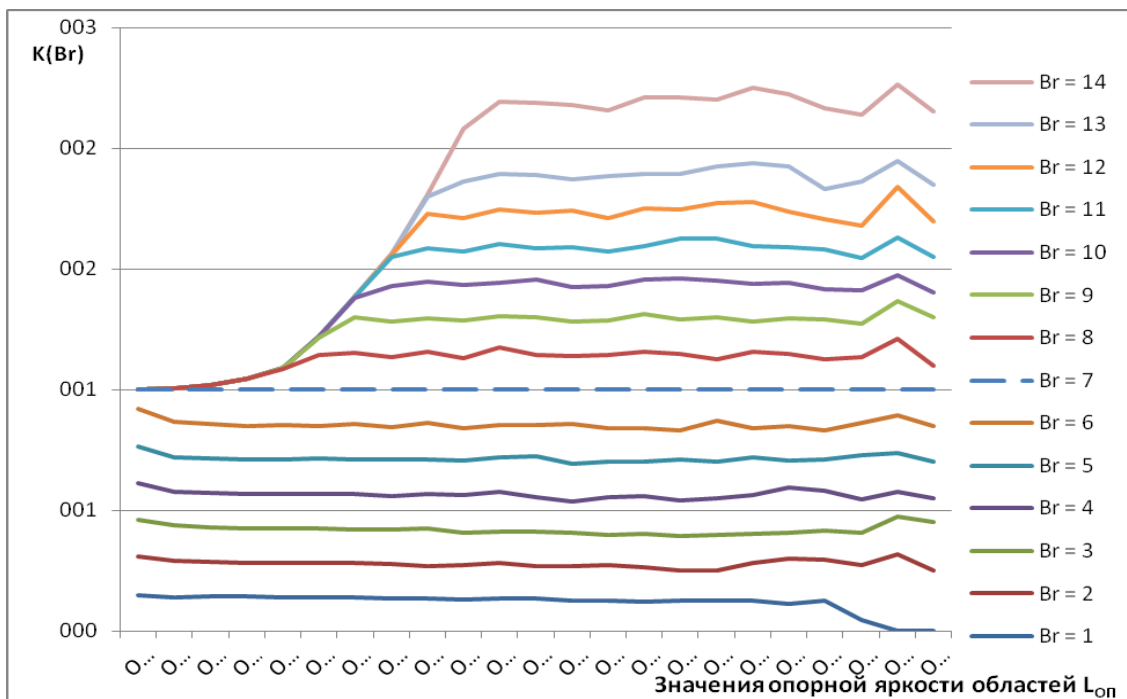


Рис. 10

Из рисунка очевидно, что при изменении уровня освещённости коэффициенты $K(Br)$ для всех областей практически неизменны. Если исходить из того, что уровень освещённости одинаков для всего изображения, то и коэффициенты изменения яркости должны быть одинаковыми. В этом случае можно установить причину различий параметра Value для ЭО – иные условия освещённости, либо различие уровней яркости объекта в данной области.

Далее необходимо проверить справедливость этого правила для цветных областей. Другими словами, необходимо выяснить, влияет ли цветовой тон на полученную выше зависимость яркости от освещённости. Для этого были получены изображения цветных областей при последовательном увеличении уровня освещённости. На рисунке 11 представлены графики изменения координаты Value для трёх цветных областей. При расчёте коэффициентов отношения уровней яркости были получены аналогичные результаты: диапазон значений оказался очень узким. Например, при повышении уровня освещённости с 2 до 7 яркость каждой из областей выросла в 3.5 раза.

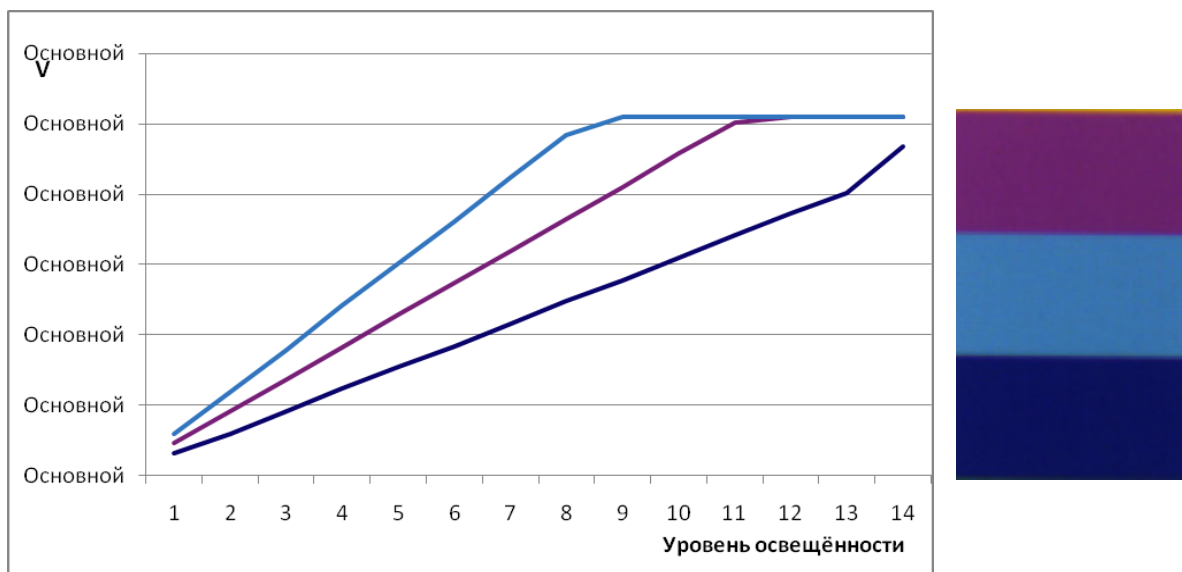


Рис. 11

На рисунках 12 и 13 показаны гистограммы яркости данных областей, снятых при уровнях освещённости 2 и 7 соответственно.

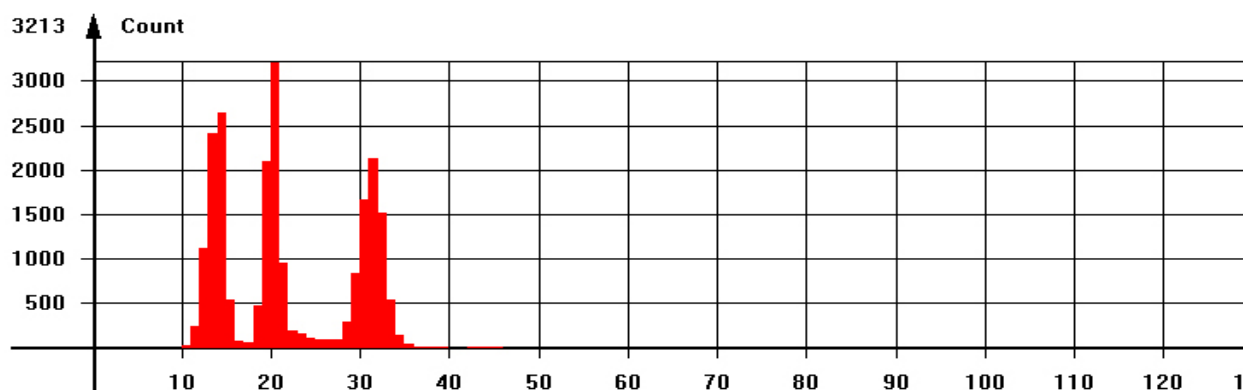


Рис. 12

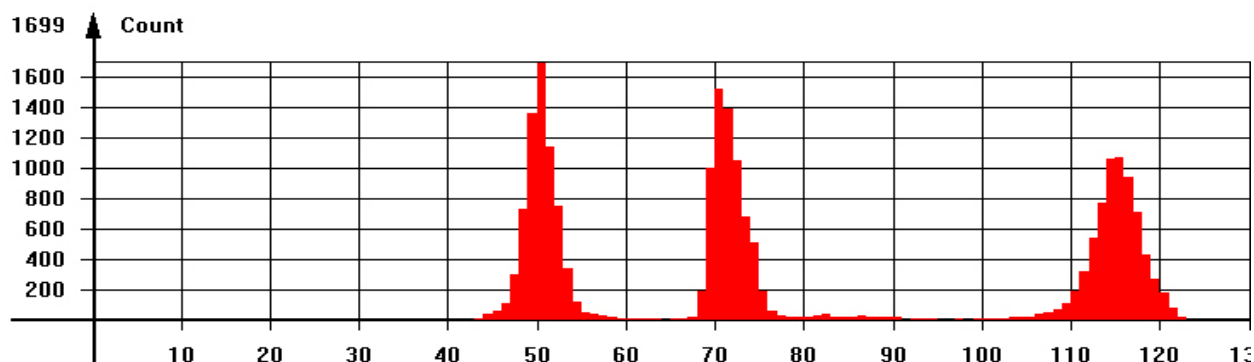


Рис. 13

Данные гистограммы подтверждают достоверность полученных закономерностей, что позволяет сделать следующие выводы:

зависимость уровня яркости от освещённости имеет линейный характер вплоть до достижения уровня насыщения;

цветовой тон не влияет на зависимость яркости от освещённости.

Следовательно, для принятия первичного решения о схожести объектов необходимо рассчитать коэффициенты $K(Br)$, показывающие изменение уровня средней яркости ЭО изображения. Если сравниваемые изображения схожи, то значения коэффициентов будут попадать в установленный доверительный интервал. Физический смысл операций расчёта и анализа значений коэффициентов: подтверждение совпадения пространственного распределения яркости вне зависимости от условий освещённости.

Порядок сравнения объектов (сцен)

Был разработан следующий алгоритм сравнения цветовых параметров объектов(сцен), позволяющий выдавать процент их схожести. Исходными данными являются изображения, а так же пороги и доверительные интервалы для параметров описания. Результатом работы является степень совпадения данных изображений по цвету. Общая схема алгоритма представлена на рисунке 14.

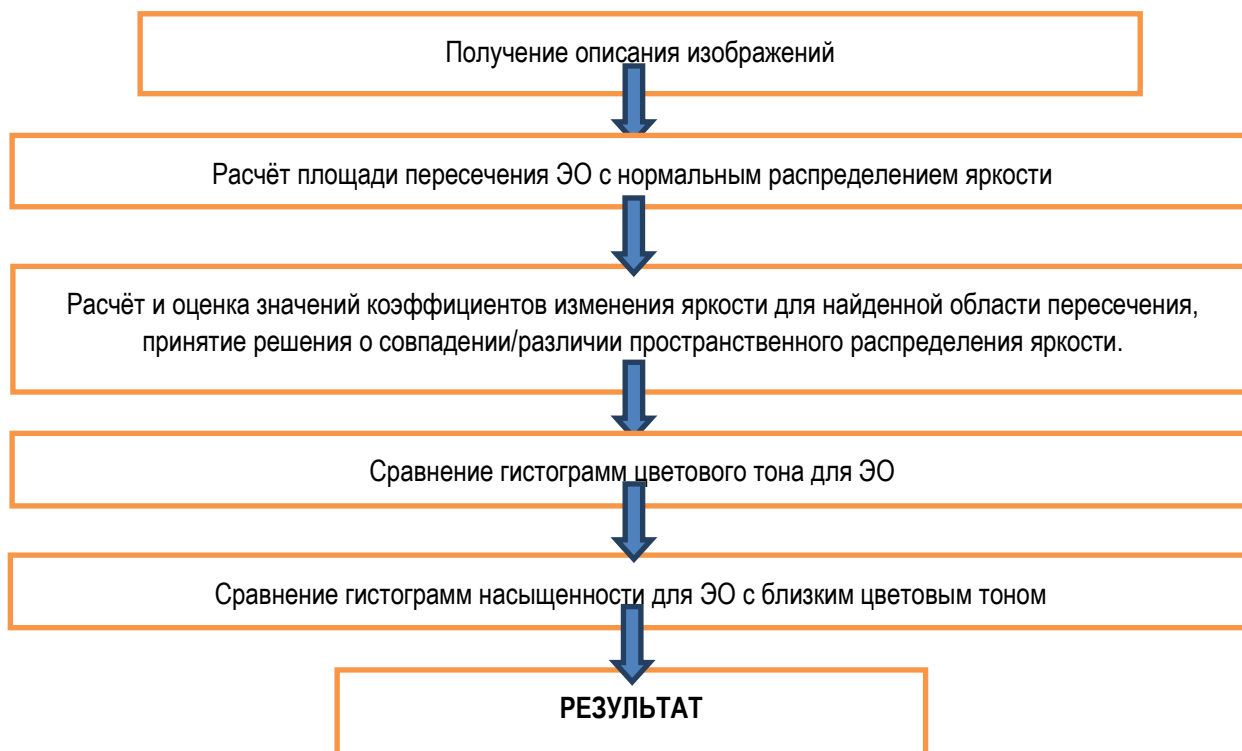


Рис. 14

Сравнение производится поэтапно с возможностью отказа от продолжения. Для каждого этапа задаётся минимально-допустимое значение степени совпадения. В случае, если полученное на одном из этапов значение ниже заданного, то проведение дальнейшего сравнения считается нецелесообразным.

Первоначально получают описания изображений с учётом указанных доверительных интервалов. После этого оценивается, насколько совпадают области с нормальным распределением яркости на рассматриваемых изображениях, рассчитывается площадь совпадения. Дальнейшая оценка проводится лишь в том случае, если объём пересечения превышает установленное пользователем значение.

Следующим этапом является расчёт коэффициентов изменения средней яркости ЭО для найденной области пересечения. Было экспериментально доказано, что большой разброс значений этих коэффициентов указывает на то, что на сравнение подаются различные объекты (сцены). Это позволяет уже на первом этапе принять решение о различии изображений из-за другого пространственного распределения яркости. В случае, если коэффициенты находятся в пределах доверительного интервала, установленного пользователем, считается что различие по яркости обусловлено лишь изменениями условий освещённости.

Далее производится переход к сравнению по каналам цветового тона и насыщенности в ЭО. Сравнение гистограмм внутри ЭО начинается со сравнения цветового тона. Минимальная площадь пересечения по цветовому тону так же задаётся пользователем. Результатом является процент площади изображения, в которых цветовой тон совпадает.

Затем производится расчёт гистограммы насыщенности, и снова производится сравнение путём нахождения пересечения. Учитывая, что до этого были получены данные о том, что значения яркости и цветового тона совпадают, то результат пересечения на данном этапе и будет служить общим результатом сравнения.

Заключение

Данный алгоритм был разработан для использования в системе распознавания документов. Полный цикл сравнения изображений, полученных с камеры с разрешением 1.3 мегапикселя, происходит менее чем за

1 миллисекунду. Благодаря ступенчатой структуре алгоритма время, которое тратится на этап распознавания по цвету, составляет порядка 300 миллисекунд для 1000 классов.

Использование свойства независимости цветового тона от уровня освещённости, а также подтверждение пространственного распределения яркости позволили решить задачу распознавания типа документа без применения структурных алгоритмов.

Библиография

[ГонсалесВудс, 2006] Р. Гонсалес, Р. Вудс «Цифровая обработка изображений». Москва: Техносфера, 2006 – 1072с.

[Fairchild, 2005]Mark D. Fairchild. Color Appearance Models, 2nd Edition. JohnWiley&Sons, 2005. – 408 p.

[Прохоров, 1988] А.М. Прохоров. Физическая энциклопедия. Москва, Советская энциклопедия, 1988, т.5. - 420с.

<http://ru.wikipedia.org/wiki>

Сведения об авторах



Протосавицкий Анатолий – инженер-программист, ООО «Регула», 220036 ул.

Волоха 1-314, г. Минск, Беларусь, e-mail: anatol.pratasavitski@regula.by

Основные направления деятельности: обработка цветных изображений, элементы защиты и системы проверки подлинности документов



Горшкова Екатерина – инженер-программист, ООО «Регула», 220036 ул. Волоха 1-

314, г. Минск, Беларусь, e-mail: katsiaryna.harshkova@regula.by

Основные направления деятельности: обработка цветных изображений, элементы защиты и системы проверки подлинности документов



Пилюттик Дмитрий – инженер-программист, ООО «Регула», 220036 ул. Волоха 1-

314, г. Минск, Беларусь, e-mail: dzmitry.piliutsk@regula.by

Основные направления деятельности: обработка изображений, системы распознавания, системы OCR.



Телятников Роман – к.т.н., научный руководитель разработками ПО, ООО

«Регула», 220036 ул. Волоха 1-314, г. Минск, Беларусь, e-mail:

raman.tsialiatnikau@regula.by

Основные направления деятельности: распознавание образов, обработка изображений, нейрофизиология



Шумский Иван– к.т.н., генеральный директор ООО «Регула», 220036 ул. Волоха 1-314, г. Минск, Беларусь, e-mail: ivan.shumsky@regula.by

Основные направления деятельности: автоматизация анализа подлинности документов и денежных банкнот: проектирование оборудования и программного обеспечения

ОНЛАЙН РАСПОЗНАВАНИЕ РУКОПИСНЫХ МАТЕМАТИЧЕСКИХ ФОРМУЛ

Надеран Эдрис

Аннотация: рассматривается разработанный алгоритм структурного анализа для онлайн распознавания рукописных математических формул. Вводятся следующие понятия: допустимое и обязательное расположение символа, алгоритм размещения, алгоритм чередования, алгоритм группировки.

Ключевые слова: структурный анализ, математические формулы, распознавание рукописных символов, онлайн распознавание, сегментация.

ACM Classification Keywords: I.5.2 Computing Methodologies - PATTERN RECOGNITION - Design Methodology - Pattern analysis. G.4 Mathematics of Computing – MATHEMATICAL SOFTWARE - Algorithm design and analysis. I.5.1 Computing Methodologies - PATTERN RECOGNITION - Models - Structural.

Вступление

Задачу распознавания математических формул можно разделить на два основных этапа: распознавание символов и структурный анализ. Такое разделение на подзадачи имеет ряд преимуществ, в частности подзадачи могут быть решены независимо друг от друга, что позволяет проанализировать результаты и сделать улучшения на каждом из этапов, при этом сохранится целостность всей системы. Целью данной статьи является подробно рассмотреть разработанный метод структурного анализа, который позволяет определить пространственные отношения между символами и полностью отобразить структуру, написанной пользователем формулы. Предлагаемый алгоритм является гибким к внесению изменений, поскольку пересматривает формулу после каждого внесенного изменения, что обеспечивает дополнительный комфорт для пользователя при введении формулы онлайн. Работоспособность алгоритма была протестирована с помощью разработанного редактора форм. Полученные результаты экспериментов приводятся в третьем разделе статьи.

1 Краткий обзор процесса распознавания

Дадим определение термина математическая формула - это символическая запись законченного логического суждения, которая может состоять из цифр, букв и различного вида символов, при этом все цифры, буквы и символы формируют внутреннюю иерархическую структуру. Следуя из этого определения можно говорить о том, что для того, чтобы правильно распознать формулу необходимо точно знать значение каждого из символов, которые вошли в формулу и определить расположение этих символов относительно друг друга.

В данной статье будет подробно рассмотрен алгоритм структурного анализа, позволяющий решить вторую из поставленных задач, а именно определить расположение символов относительно друг друга. Но сначала кратко рассмотрим процесс сегментации и непосредственно распознавания каждого отдельно сегментированного символа, и после перейдем к описанию самого метода структурного анализа. Такой подход позволит создать полную картину процесса онлайн распознавания рукописных формул. Необходимо заметить, что все три этапа: сегментация, распознавание и структурный анализ происходят друг за другом для каждого отдельно сегментированного образца.

В предлагаемой системе, используется набор из 77 символов, который позволяет писать тригонометрические, логарифмические функции, интегралы, греческие и латинские буквы. В работе

рассматривается система онлайн распознавания, как наиболее перспективная область развития в ближайшие несколько лет, по причине увеличения использования планшетных ПК.

0	7	e	l	s	z	{	.	Ω	$\sqrt{\quad}$	Δ
1	8	f	m	t	+	}	,	π	!	∇
2	9	g	n	u	-	[Σ	α	\approx	λ
3	a	h	o	v	/]	∞	β	γ	σ
4	b	i	p	w	*	<	\geq	μ	θ	\int
5	c	j	q	x	(>	\leq	φ	ε	∂
6	d	k	r	y)	%	=	δ	\in	\exists

Рис. 1 Рассматриваемые символы в системе онлайн распознавания формул

На первом этапе происходит сегментация символов. В рукописном математическом выражении символы почти всегда печатные, не слитные. То есть каждый символ написан отдельно от всех остальных. На этом предположении основывается алгоритм разделения. Необходимо отметить, что при добавлении новых штрихов или удалении уже существующих, может значительно измениться значение символа, поэтому происходит динамическая интерпретация символа после каждого внесенного пользователем изменения.

Для решения задачи распознавания изолированных символов был выбран нечеткий классификатор NEFClass. Система NEFClass прекрасно справляется с необходимой задачей классификации символов, поскольку сочетает в себе нейронные сети и нечеткие системы логического вывода, при этом обладает высокой скоростью работы алгоритмов обучения ([1], [2]). Для задачи распознавания рукописных символов был построен нечеткий классификатор с 3 входами и 77 выходами. Входам системы соответствуют признаки, которые были подобраны специально для математических символов, цифр, латинских и греческих букв, а выходам соответствует набор из 77 символов, которые пользователь может использовать при написании формул. В качестве алгоритма обучения был выбран алгоритм сопряженных градиентов, который обеспечивает сходимость к оптимальному решению за конечное число шагов ([1], [2]). В результате этапа распознавания, на выходе нейронной сети будет получен один из символов.

2 Структурный анализ

Основной задачей этапа структурного анализа является определить расположение символов в формуле относительно друг друга.

В разработанном методе структурного анализа было выделено несколько основных понятий:

- Допустимое и обязательное расположение
- Алгоритм размещения
- Алгоритм чередования
- Алгоритм группировки

Рассмотрим предлагаемый метод более подробно.

Все символы были разделены на группы, в зависимости от допустимых позиций размещения других символов относительно них. На Рисунке 2 схематически изображены все восемь возможных позиций.



Рис. 2 Схематическое изображение допустимых позиций расположения символов относительно текущего рассматриваемого символа

Для разработанного метода были выделены следующие группы символов:

1. Цифры. Принимают значения от 0 до 9, π , ∞ . Символы могут быть размещены относительно данной группы в следующих позициях – 1, 2.

Математические знаки. К этой группе были отнесены такие символы: *, +, -, /, =, <, ≤, >, ≥, %, ≈, €, ∃, Δ,

∇. Допустимо размещение символов во 2 позиции.

2. Скобки открывающие. В эту группу вошли следующие символы: (, {, [. Относительно этой группы символы могут располагаться во 2 позиции.
3. Скобки закрывающие. К этой группе относятся такие символы:), },]. Допустимое размещение символов относительно рассматриваемого символа: 1, 2 позиции.
4. Знаки препинания. К знакам препинания были отнесены: «.», «,», «!». Символы могут располагаться в 1, 2 позиции.

Интеграл \int . Допустимое размещение символов во 2, 4, 8 позиции.

5. Извлечение корня $\sqrt{\quad}$. Допустимое размещение: 1, 2, 6 позиции.
6. Сумма \sum . Допустимое расположение символов: 2, 4, 8 позиции.
7. Греческие и латинские буквы, рассмотренные на Рис.1. Допустимое размещение символов: 1, 2, 3 позиции.
8. Тригонометрические функции. Рассматриваются следующие тригонометрические функции \sin , \cos , \tan/tg , $\cotg/ctn/cot$. Допустимое размещение символов: 1, 2 позиции.
9. Логарифмические функции \ln , \lg , \log . Допустимое размещение символов: 1, 2, 3 позиции.

Под допустимой позицией расположения символов понимается возможная, но не обязательная к заполнению символом позиция. Под термином «обязательная позиция» имеется в виду такая позиция, которая должна быть обязательно заполнена одним из следующих символов, такая позиция получит приоритет при добавлении нового символа. Например, знак \int или $\sqrt{\quad}$ предполагает обязательное наличие зависимого символа в позиции 2.

Алгоритм размещения нового символа относительно других символов работает по следующему принципу.

1. Формула перезаписывается слева направо каждый раз при добавлении нового символа.
2. Выбирается позиция для нового символа А по следующему правилу:
 - Проверяется наличие пустой обязательной позиции, в случае ее наличия данная позиция получает приоритет и в нее помещается символ А.
 - Если пустая обязательная позиция отсутствует, то выбирается ближайшая допустимая позиция, в которую помещается новый символ А.
3. В случае, если позиция, в которую был помещен новый символ А, была занята символом Б, то выбирается новая позиция для символа Б по правилам описанным в пункте 2.

Стоит отдельно рассмотреть знаки «.» и «-». Так точка в выражении может быть десятичной запятой или оператором умножения в зависимости от положения точки и ее соседних символов. Горизонтальная линия может быть дробной чертой или знаком минус в зависимости от длины линии и наличия символов выше и ниже линии. В предлагаемом методе для решения для решения такого рода неопределенности были использованы следующие подходы:

- Знак «-» будет рассматриваться как дробная черта, в том случае, если имеет символы в 4 и/или 8 позициях.
- Знак «.» рассматривается как оператор умножения, если центр масс его сегмента находится ближе к 2 позиции соседнего слева символа.

Несколько букв написанных вместе могут формировать единое значение. Это могут быть тригонометрические функции \sin , \cos , \tan/tg , $ctg/ctn/cot$ или логарифмические функции \ln , \lg , \log . Рассматривая группу букв написанных подряд алгоритм проверяет, не образуют ли они название функции. В этом случае происходит чередование нескольких букв на одну из известных функций. Кроме того, алгоритм делает проверку различных вариантов, так, если на этапе распознавания одна из букв была распознана неправильно, на данном этапе ошибка будет исправлена. На Рис. 3 приведен пример работы описанного алгоритма чередования:

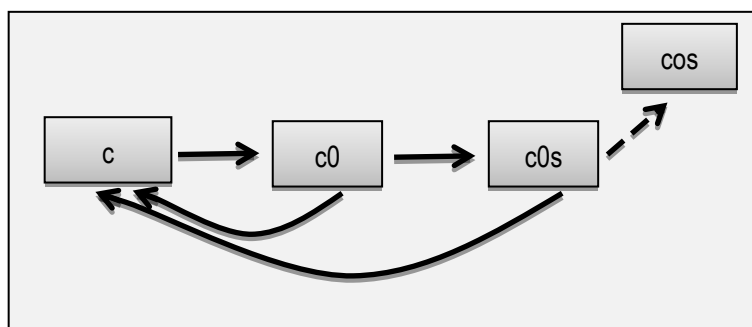


Рис. 3 Пример работы алгоритма чередования

На Рис.3 видно, что каждый раз при добавлении нового символа формула начинает просматриваться сначала, в данном случае буква «o» на этапе распознавания была ошибочно распознана как цифра «0». После добавления следующего символа алгоритм просматривает все возможные комбинации, имеющиеся в базе, и, найдя наиболее подходящую, исправит ошибку.

Следующее понятие, которое было введено в разработанном методе – это понятие группы. Специальные символы, к ним были отнесены различного вида скобки, дробная черта и знаки препинания: точка, запятая, позволяют сгруппировать несколько отдельных значений в одну группу. Алгоритм группировки позволяет корректировать длину скобок и дробной черты, выделить дробное число, как результат такой группировки соответствующая оригиналу формула будет отображена наиболее верно.

Использование описанного метода позволяет определить пространственные отношения между символами и полностью отобразить структуру, написанной пользователем формулы. Кроме, того полный анализ структуры формулы после добавления каждого нового символа делает систему гибкой к внесению изменений.

3 Экспериментальные исследования

Для проверки эффективности работы рассматриваемого метода структурного анализа были проведены тесты в разработанном редакторе формул. Для тестирования были отобраны 50 различных формул, включающих в себя тригонометрические, логарифмические функции, операции возведения в степень, извлечения корня, дробного деления, индексы. На Рис. 4 видно, что редактор хорошо структурирует достаточно сложную формулу с использованием тригонометрических функций.

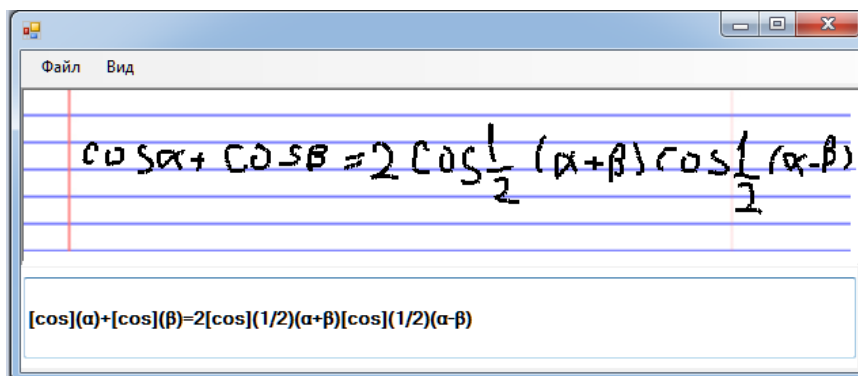


Рис. 4 Пример работы редактора формул

В таблице 1 приводятся результаты работы алгоритма по количеству неправильно структурированных символов. Можно сказать, что такие символы как буквы и знаки операций имеют меньший процент ошибки, в отличие от логарифмических функций и знака суммы.

Таблица 1. Результаты работы алгоритма структурного анализа

	Кол-во символов, содержащихся в протестированных формулах	Кол-во неправильно структурированных символов	% неправильно структурированных символов
Цифры	371	14	6,46
Знаки операций	194	10	5,15
Скобки	92	8	8,69
Знаки препинания	17	3	17,6
Интеграл	25	3	12
Извлечение корня	23	2	8,69
Сумма	15	2	13,3
Греческие и/или латинские буквы	160	7	4,37
Тригонометрические функции	58	4	6,89
Логарифмические функции	15	2	13,3

На Рис. 5 приводится пример неправильно определенной структуры формулы. Здесь алгоритм структурного анализа неправильно определил пространственные отношения между символами. Это ошибка могла произойти по причине слишком близкого расположения значения «2» к полю обязательному полю тригонометрической функции sin.

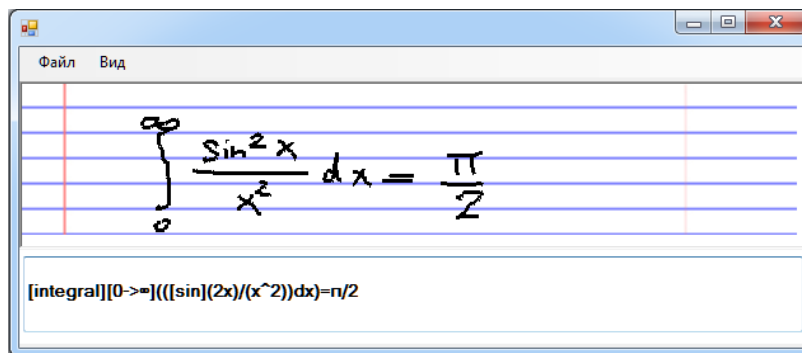


Рис. 5 Пример работы редактора формул

В результате тестирования было получено общее значение ошибочно классифицированных символов 6,7% с проверочного набора данных. Данное значение свидетельствует о достаточной эффективности использования предложенного в работе метода структурного анализа к онлайн распознаванию рукописных формул.

Выводы и перспективы дальнейших исследований

В результате проведенного исследования был разработан алгоритм структурного анализа, который позволяет достаточно точно определить пространственные отношения между символами и как результат полностью отобразить структуру, написанной пользователем формулы. Общее значение ошибочно классифицированных символов равно 6,7% свидетельствует об эффективности его применения к поставленной задаче. Как показали исследования такие символы как логарифмические функции, знак суммы, знаки препинания имеют самый большой процент неправильно структурированных символов. Это можно объяснить наличием нескольких обязательных позиций у данных символов. Необходимо провести дальнейшие исследования, которые позволят улучшить алгоритм. Улучшения будут достигнуты путем доработки алгоритма размещения. На данном этапе модель алгоритма структурного анализа разработана с учетом правила, что пользователь может ввести только одну формулу в редакторе. Для того, чтобы пользователь мог ввести нескольких формул одновременно алгоритм должен быть усовершенствован, введением дополнительных символов, которые будут разделять формулы. Кроме того необходимо разработать систему подсказок, которая позволит пользователю выбирать одну из предлагаемых в списке структур. Предлагаемые улучшения позволят еще больше снизить процент ошибочно классифицированных символов.

Литература

1. Зайченко Ю.П. Нечёткие модели и методы в интеллектуальных системах. Учебное пособие для студентов высших учебных заведений. – К.: «Издательский Дом «Слово»», 2008. – С. 344
2. Зайченко Ю.П. Основы проектирования интеллектуальных систем. Учебное пособие. – К.: «Издательский Дом «Слово»», 2004. – С. 352

Информация про автора



Надеран Эдрис – аспирант Национального технического университета Украины «КПИ», адрес электронной почты: e.naderan@gmail.com

Основные сферы научных исследований автора: применение нечеткого классификатора к онлайн распознаванию рукописных формул.

ЛИНЕЙНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ МОДЕЛЕЙ ФИНАНСОВЫХ ВРЕМЕННЫХ РЯДОВ

Зайченко Ю.П., Мурга Н.А.

Аннотация: предлагается метод линейного оценивания параметров моделей финансовых временных рядов противоположный по своей структуре МНК. Проводится сравнительный анализ МНК и предлагаемого метода.

Ключевые слова: линейное оценивание, оптимизация, финансовые временные ряды.

Ключевые слова классификации ACM: G. Mathematics of Computing - G.1 NUMERICAL ANALYSIS - G.1.2 Approximation - Linear approximation.

Вступление

В работе рассматривается метод оценивания параметров линейной модели финансовых временных рядов противоположный по критерию оптимизации МНК. Если МНК восстанавливает параметры модели, минимизируя сумму квадратов отклонений реальных значений от выдаваемых моделью, то рассматриваемый метод минимизирует максимальное отклонение. Актуальность данной тематики обусловлена тем, что трейдер принимает решение не «в среднем», а в конкретные моменты времени и для него важно именно то решение и тот прогноз, которые он принимает в данный конкретный момент времени, а не прогноз «в среднем». Опора на среднеквадратическую оценку параметров модели не позволяет оценить возможность ситуации маржин-колл, что является неприемлемым при реальной торговле. Кроме вышесказанного, предлагаемый метод не имеет ограничений МНК: некоррелируемость входов модели, входов модели и шумов и нулевое математическое ожидание шумов. Всё это обусловило необходимость разработки предлагаемого метода.

Описание метода

Пусть x_1, x_2, \dots, x_n - наблюдаемые значения переменных; y - наблюдаемое значение, которое, предполагается, линейно зависит от x_1, x_2, \dots, x_n . Пусть $x_{1i}, x_{2i}, \dots, x_{ni}$ и y_i , где $i = 1 \dots L$ - конкретные значения этих переменных. Линейная зависимость описывается формулой (1):

$$y = \sum_{j=1}^n a_j x_j \quad (1)$$

Если ввести обозначение $d_i = y_i - \sum_{j=1}^n a_j x_{ij}$, то задача восстановления функциональной зависимости сводится к нахождению таких $a_j, j = 1 \dots n$, чтобы $d_i = 0, i = 1 \dots L$. Очевидно, что при применении к реальным задачам, в частности – к задаче анализа данных финансовых рынков, данное равенство не будет иметь место, так как имеются шумы, искажающие данные, выборка может быть сформирована неверно (неправильно выбрано n) и т.д.

Классический МНК предполагает решение задачи безусловной оптимизации (2):

$$\sum_{i=1}^L d_i^2 \rightarrow \min \quad (2)$$

Однако решение этой задачи оптимизации имеет один существенный (по крайней мере, для торговли на финансовых рынках недостаток): в результате решения задачи все d_i , кроме маленькой группы могут оказаться малыми, а в этой группе – большими, то есть метод пожертвовал точностью для малой группы отклонений в пользу суммарной ошибки на всей выборке. На практике, например биржевой торговли, это говорит о том, что есть некоторая вероятность открыть убыточную позицию. С целью устранения этого недостатка предлагается рассматриваемый ниже метод.

Прежде всего, необходимо ввести понятие нечёткого нуля.

Пусть нечётким нулём будет называться нечёткое множество O , на котором определена функция принадлежности данному множеству (3):

$$\mu_O(z) = \exp\left\{-\frac{z^2}{\sigma^2}\right\} \tag{3}$$

График данной функции изображён на Рисунке 1.

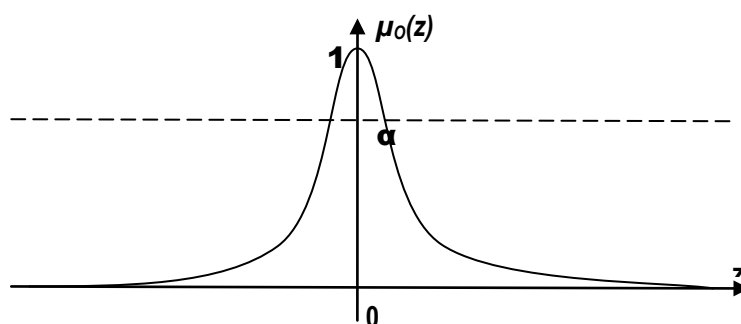


Рисунок 1. График функции принадлежности множеству «нечёткий нуль»

Суть предлагаемого метода состоит в нахождении таких $d_i, i = 1 \dots L$, чтобы $\mu_O(d_i) \rightarrow \max \forall i = 1 \dots L$. Если зафиксировать некоторый уровень значимости α , на котором будет происходить оптимизация, то задача оптимизации может быть сведена к следующей:

$$\begin{aligned} &\min \sigma, \\ &d_i \leq \sigma \sqrt{\ln \frac{1}{\alpha}}, i = 1 \dots L, \\ &d_i \geq -\sigma \sqrt{\ln \frac{1}{\alpha}}, i = 1 \dots L, \\ &\sigma \geq 0. \end{aligned} \tag{4}$$

Если, согласно введённым ранее обозначениям, расписать d_i , то задача (4) преобразуется к виду (5):

$$\begin{aligned} &\min \sigma, \\ &y_i - \sum_{j=1}^n a_j x_{ij} \leq \sigma \sqrt{\ln \frac{1}{\alpha}}, i = 1 \dots L, \\ &y_i - \sum_{j=1}^n a_j x_{ij} \geq -\sigma \sqrt{\ln \frac{1}{\alpha}}, i = 1 \dots L, \\ &\sigma \geq 0. \end{aligned} \tag{5}$$

Если каждый a_j переписать в виде $a_j = a_{j1} - a_{j2}$, $a_{j1}, a_{j2} \geq 0$ и y_i перенести в правую часть неравенств, а $\pm \sigma \sqrt{\ln \frac{1}{\alpha}}$ в левую, то можно получить следующую задачу линейного программирования:

$$\begin{aligned}
& \min \sigma, \\
& y_i \leq \sigma \sqrt{\ln \frac{1}{\alpha}} + \sum_{j=1}^n a_{j1} x_{ij} - \sum_{j=1}^n a_{j2} x_{ij}, \quad i=1 \dots L, \\
& -y_i \leq \sigma \sqrt{\ln \frac{1}{\alpha}} - \sum_{j=1}^n a_{j1} x_{ij} + \sum_{j=1}^n a_{j2} x_{ij}, \quad i=1 \dots L, \\
& \sigma \geq 0, a_{j1}, a_{j2} \geq 0 \quad j=1 \dots n.
\end{aligned} \tag{6}$$

Так как на практике L достаточно велико, то имеет смысл (с целью минимизации расходов вычислительных ресурсов) перейти от (6) к двойственной задаче (7) и решать именно её. Пусть $z_k, k=1 \dots 2L$ - двойственные переменные, тогда двойственная задача к задаче (6) имеет вид:

$$\begin{aligned}
& \max \sum_{i=1}^L (z_i y_i - z_{L+i} y_i), \\
& \sum_{i=1}^{2L} z_i = \frac{1}{\sqrt{\ln \frac{1}{\alpha}}}, \\
& \sum_{i=1}^L (x_{ij} z_i - x_{ij} z_{i+L}) \leq 0, \quad j=1 \dots n, \\
& \sum_{i=1}^L (-x_{ij} z_i + x_{ij} z_{i+L}) \leq 0, \quad j=1 \dots n, \\
& z_i \geq 0, \quad i=1 \dots 2L.
\end{aligned} \tag{7}$$

Наиболее экономично решать эту задачу методом обратной матрицы, и получив решение, восстановить значения $\sigma \geq 0, a_{j1}, a_{j2} \geq 0 \quad j=1 \dots n$ по оценкам метода. Данный метод не допускает возможности пренебречь качеством отдельных значений выборки в пользу минимизации общего критерия и, также, позволяет получить сразу диапазон точности оценки коэффициентов.

Классификация методов линейного оценивания

Сразу следует оговориться, что о задаче фильтрации в данной части речи идти не будет. Если читатель захочет ознакомиться с методами фильтрации, то его можно отправить к замечательным работам [1;2;3].

Первый признак, по которому будет идти классификация, – это задача оптимизации, которая будет решаться. Так в МНК оценивании, как было сказано раньше, решается задача безусловной минимизации суммы квадратов отклонений, которые даёт модель от желаемых значений. Метод же, предложенный в предыдущей части, производит минимизацию максимального отклонения. Однако это же выполняет ещё один метод, описанный в [4].

Особенность данного метода состоит в том, что a_j в (1) – это нечёткие числа, которые выбраны в работе [4] треугольного, гауссовского и колоколообразного вида. Для случая гауссовского вида каждое a_j задано центром функции принадлежности и «рассевом» - (c_j, σ_j) , и, таким образом, в формуле (3) параметр σ определяется соотношением (8):

$$\sigma = \sum_{j=1}^n \sigma_j x_j \tag{8}$$

Таким образом, задача (6) в данном случае может быть переписана в виде (9):

$$\begin{aligned}
& \min \sum_{i=1}^L \sigma_i, \\
& y_i \leq \sigma_i \sqrt{\ln \frac{1}{\alpha}} + \sum_{j=1}^n a_{j1} x_{ij} - \sum_{j=1}^n a_{j2} x_{ij}, \quad i=1 \dots L, \\
& -y_i \leq \sigma_i \sqrt{\ln \frac{1}{\alpha}} - \sum_{j=1}^n a_{j1} x_{ij} + \sum_{j=1}^n a_{j2} x_{ij}, \quad i=1 \dots L, \\
& \sigma_i \geq 0, a_{j1}, a_{j2} \geq 0 \quad j=1 \dots n,
\end{aligned} \tag{9}$$

где $\sigma_i = \sum_{j=1}^n \sigma_j x_{ij}$.

Таким образом, данные два метода ((6) и (9)) могут быть классифицированы по виду модели оценивания: чёткой либо нечёткой, и, соответственно, параметр σ в формуле (3) будет либо константой – метод (6) – либо линейной комбинацией входов модели – метод (9).

Интересным представляется вопрос о связи метода (6) и метода (9).

Если выбрать число σ согласно с формулой (10):

$$\sigma = \max_i \sigma_i \tag{10}$$

то очевидным является тот факт, что

$$y_i \leq \sigma_i \sqrt{\ln \frac{1}{\alpha}} + \sum_{j=1}^n a_{j1} x_{ij} - \sum_{j=1}^n a_{j2} x_{ij} \leq \sigma \sqrt{\ln \frac{1}{\alpha}} + \sum_{j=1}^n a_{j1} x_{ij} - \sum_{j=1}^n a_{j2} x_{ij} \tag{11}$$

и

$$-y_i \leq \sigma_i \sqrt{\ln \frac{1}{\alpha}} - \sum_{j=1}^n a_{j1} x_{ij} + \sum_{j=1}^n a_{j2} x_{ij} \leq \sigma \sqrt{\ln \frac{1}{\alpha}} - \sum_{j=1}^n a_{j1} x_{ij} + \sum_{j=1}^n a_{j2} x_{ij} \tag{12}$$

И задача линейного оценивания коэффициентов модели может быть переписана в виде (6).

Таким образом, исходя из приведённых выше выкладок, можно заключить, что решение задачи (9) даёт не худшие оценки коэффициентов модели, чем задачи (6). Кроме того, следует заметить, что аналогичный задаче (6) коэффициент отклонения результата для задачи (9) может быть оценён формулой (10).

В применении всех описанных выше методов к анализу финансовых рынков есть свои преимущества и недостатки.

МНК следует применять, если идёт торговля на выигрыш «в среднем». Однако, следует отметить, что при такой торговле весьма вероятна ситуация маржин-колл, в следствие указанного ранее очень сильного отклонения прогноза модели от реального значения котировки. Кроме того, при МНК оценке необходимо быть уверенным в том, что математическое ожидание шумов близко к нулю, нет корреляции между шумами и входами модели, между входами модели (а она есть, так как рассматривается временной ряд)... Возникают проблемы с обращением матриц.

Предложенный в работе метод избавлен от этих недостатков: инвестор получает предел отклонения прогноза от реальных значений (то есть реально оценивает вероятность ситуации маржин-колл), всевозможные нелинейности относительно шумов уже заложены в модель, оценивание происходит путём решения задачи линейного программирования, которая для данной задачи обязательно имеет хотя бы единственное решение.

Метод (9) решает аналогичную методу (6) задачу, но в среднем даёт более уточнённую оценку параметров модели. Однако, исходя из вида критерия оптимизации, данный метод, как и МНК ищет модель «в среднем». Но данный метод, как и метод (6) избавлен от тех же недостатков МНК, как и (6).

Практическое исследование метода

Для проверки теоретических выкладок, предложенных в данной работе, проводились исследования линейного оценивания функциональных зависимостей временных рядов котировок валютных пар EUR/USD, EUR/GBP, USD/JPY, USD/CHF, взятых из [5], котировок мировых финансовых индексов CAC40, DAX, DJIA, HANG SENG, NASDAQ, MMBБ, RTS2 и финансовых инструментов COMEX GOLD, MMBБ GSZP, ICE.BRN, которые были взяты из [6]. Для проведения опытов выборки данных финансовых инструментов разбивались на обучающие и проверочные в соотношениях соответственно 50%:50%. В сравнительном анализе участвовали: МНК, с градиентным методом восстановления значений параметров и метод (6).

Если x_i - значение котировки в момент времени i , то преобразованное значение котировки определяется формулой (13):

$$y_i = \frac{x_i - \min_i x_i}{\max_i x_i - \min_i x_i} \quad (13)$$

область определения преобразованной котировки – интервал $[0;1]$.

Показатели эффективности работы методов на собственно котировках x_i - это критерии *MAPE* (14) и *max APE* (15):

$$MAPE = \frac{1}{L} \sum_{i=1}^L \frac{|x_i - \bar{x}_i|}{x_i} \quad (14)$$

$$\max APE = \max_i \frac{|x_i - \bar{x}_i|}{x_i} \quad (15)$$

где x_i - реальное значение котировки, а \bar{x}_i - значение котировки, спрогнозированное моделью.

Показатели эффективности работы методов на модифицированных котировках y_i - это критерии *RMSE* (16) и *max Error* (17):

$$RMSE = \sqrt{\frac{1}{L} \sum_{i=1}^L (y_i - \bar{y}_i)^2} \quad (16)$$

$$\max Error = \max_i |y_i - \bar{y}_i| \quad (17)$$

При проведении экспериментов для всех котировок было установлено, что на качество оценивания выбор уровня значимости α не влияет как при анализе временного ряда x_i , так и при анализе y_i . Однако для более стабильной работы метода (6) лучше всего выбирать α в диапазоне $[0,5;0,7]$. Этот результат обоснован теоретически, исходя из задачи оптимизации (6) – происходит оптимизация с целью получения наименьшего σ , не зависимо от желаний проектировщика. Выбор α , как можно увидеть из первого ограничения задачи (7), влияет на величину значений оптимизируемых переменных. На Рисунке 2.

Приведён график функции $\frac{1}{\sqrt{\ln \frac{1}{\alpha}}}$, при α изменяющемся в пределах $[0,001;0,999]$, из которого

видно, что участок $[0,5;0,7]$ даёт приблизительно одинаковое значение функции - значение в пределах $[1;2]$ - что и обуславливает вычислительную эффективность участка $[0,5;0,7]$.



Рисунок 2. График функции $f(\alpha) = \frac{1}{\sqrt{\ln 1/\alpha}}$

Предварительные исследования показали, что значения котировок приведённых выше финансовых инструментов и значения мировых финансовых индексов определяются более всего предыдущим и пред-предыдущим значением соответствующего временного ряда.

Как и следовало ожидать, для зависимости котировок и индексов от двух и трёх предыдущих значений по критериям (14) и (16) эффективнее оказался МНК, а по критериям (15) и (17) – метод (6). Для прочих зависимостей временных рядов МНК и (6) отличались слабо и определялись результаты экспериментов удачностью рандомизации. Метод (9) в экспериментах участия не принимал, так как он косвенно был сравнен при сравнительных анализах МГУА и НМГУА в более ранних работах [7, 8].

В силу ограниченности объёма работы привести результаты всех экспериментальных исследований не представляется возможным. Однако для наглядности результаты сравнительного анализа МНК и (6) на значениях индекса САС40 приведены в Таблице 1.

Таблица 1. Результаты сравнительного анализа МНК и метода (6) на значениях индекса САС40

Количество периодов временного ряда, от которых зависело значение инструмента	МНК				Метод (6)			
	MAPE обуч. выборка	MAPE пров. выборка	maxAPE обуч. выборка	maxAPE пров. выборка	MAPE обуч. выборка	MAPE пров. выборка	maxAPE обуч. выборка	maxAPE пров. выборка
2	0,009683	0,011236	0,052	0,091	0,014826	0,015242	0,055	0,0695
3	0,009891	0,011139	0,042	0,086	0,020213	0,018516	0,057	0,081
4	0,010397	0,011028	0,047	0,095	0,016749	0,018113	0,052	0,074
5	0,009546	0,011431	0,044	0,083	0,012061	0,013114	0,038	0,077
6	0,010226	0,011221	0,044	0,102	0,015644	0,016394	0,049	0,062
7	0,009677	0,011295	0,045	0,091	0,017751	0,018038	0,04	0,0696
8	0,010729	0,010224	0,043	0,093	0,019501	0,021409	0,047	0,062
9	0,010203	0,010533	0,046	0,093	0,017546	0,022202	0,044	0,073

Заклучение

В работе был предложен метод линейного оценивания противоположный по своей структуре МНК. Исходя из теоретических выкладок и практических исследований, было установлено, что в среднем по выборке данных эффективнее является МНК, однако относительно каждого отдельного значения временного ряда в значимых для моделирования случаях эффективнее был метод, предложенный в работе. Кроме того, на метод не налагаются такие ограничения как на МНК: некоррелируемость значений входов модели, некоррелируемость входов модели и шумов, нулевое математическое ожидание шумов, что делает данный метод полезным средством анализа финансовых временных рядов.

Благодарности

Статья частично финансирована из проекта ITHEA XXI Института Информационных теорий и Приложений FOI ITHEA и консорциума FOI Bulgaria (www.ithea.org, www.foibg.com).

Литература

Балакришнан А.В. Теория фильтрации Калмана: пер. с англ. – М.: Мир, 1988. – 168 с.

Браммер К., Зиффлинг Г. Фильтр Калмана-Бьюси: пер. с нем. – М.: Главная редакция физико-математической литературы, 1982. – 200 с.

Синицын И.Н. Фильтры Калмана и Пугачева. – М.: Университетская книга, Логос, 2006. – 640 с.

Зайченко Ю.П. Нечёткие модели и методы в интеллектуальных системах. – К.: «Издательский Дом «Слово», 2008. – 344 с.

Дневные котировки валютных пар EUR/GBP, EUR/USD, USD/JPY, USD/CHF за период с 25.03.2009 по 24.03.2010. <http://www.finam.ru/analysis/export/default.asp>

Дневные значения мировых финансовых индексов CAC40, DAX, DJIA, HANG SENG, NASDAQ, RTS2, индекс ММББ и дневные котировки финансовых инструментов COMEX GOLD, ICE.BRN, ММББ «Газпром нефть» - GAZP за период с 1.02.2010 по 1.02.2011. <http://www.finam.ru/analysis/export/default.asp>

Зайченко Ю.П. Основы проектирования интеллектуальных систем. – К.: Видавничий Дім «Слово», 2004. – 352 с.

Зайченко Ю.П., Заєць І.О. Синтез і адаптація нечітких прогнозуючих моделей на основі методу самоорганізації // Наукові вісті НТУУ «КПІ». – 2001. - №3. – С. 34-41.

Информация про авторов

Зайченко Юрий Петрович – доктор технических наук, профессор, НТУУ «КПИ» УНК «ИПСА», адрес электронной почты: baskervil@voliacable.com

Мурга Николай Алексеевич – аспирант НТУУ «КПИ», адрес электронной почты: murqa.nicholas@gmail.com

ИССЛЕДОВАНИЕ МНОГОКРИТЕРИАЛЬНОЙ ЗАДАЧИ ОПТИМИЗАЦИИ ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ В НЕЧЕТКИХ УСЛОВИЯХ

Малихех Есфандиярфард , Юрий Зайченко, Ови Нафас Агаи Аг Гамиш

Abstract: *The problem of multicriteria fuzzy portfolio optimization is considered and investigated. This problem include two main criteria – portfolio profitability and risk . A mathematical model of this problem was constructed , explored and the sufficient conditions for its convexity were obtained. . The results of experimental investigations of solutions are presented and discussed*

Keywords: *fuzzy portfolio optimization, multicriteria optimization problem, portfolio profitability , portfolio risk*

ACM Classification Keywords: *G.1.0 Mathematics of Computing– General – Error analysis; G.1.6 Mathematics of Computing – NUMERICAL ANALYSIS – Optimization - Gradient methods, Least squares methods; I.2.3 Computing Methodologies - ARTIFICIAL INTELLIGENCE - Uncertainty, “fuzzy”, and probabilistic reasoning; I.2.6 Computing Methodologies - ARTIFICIAL INTELLIGENCE – Learning - Connectionism and neural nets;*

Introduction

В последние годы в Украине сформировалась и довольно динамично развивается новая составляющая экономики - финансовый рынок, важной частью которого является рынок ценных бумаг. С развитием этого сектора рынка возрастает актуальность финансовых вложений в ценные бумаги. Данная проблема обусловлена ее сложностью, разнообразием субъектов и объектов фондового рынка. Особенностью данной проблемы является существенная неопределенность исходной информации относительно доходности ценных бумаг (ЦБ) в будущий момент времени. Новый подход к задаче оптимизации портфеля, который позволяет учесть неопределенность исходных данных и является альтернативой классической модели Марковица, базируется на применении аппарата нечетких множеств. Проблема нечеткой портфельной оптимизации была рассмотрена и исследована в работах [[Зайченко, 2007; Зайченко, 2008]. В этих работах рассматривалась следующая постановка задачи: необходимо оптимизировать ожидаемую доходность портфеля при ограничениях на возможный риск. Алгоритм для решения этой задачи был предложен и исследован в [Зайченко, 2008] . В работе [Зайченко, Малихех, 2007] было предложено использовать прогнозирование доходностей акций, что позволило повысить эффективность получаемых решений. В работе [Зайченко, 2010] была рассмотрена двойственная задача нечеткой портфельной оптимизации -минимизация риска при ограничении на ожидаемую доходность оптимального портфеля и получены достаточные условия ее выпуклости

Целью настоящей работы является рассмотрение многокритериальной задачи нечеткой портфельной оптимизации, построение математической модели этой задачи и ее экспериментальные исследования.

Многокритериальная задача нечеткой портфельной оптимизации

Исходная задача оптимизации нечеткого портфеля, которую естественно называть прямой, имеет следующий вид: [Зайченко, 2007]

Найти ожидаемую доходность нечеткого портфеля

$$\tilde{r} = \sum_{i=1}^N \tilde{r}_i x_i \quad (1)$$

при ограничениях на риск

$$\beta(x) \leq \beta_{\text{задан}} \quad 0 < \beta < 1 \quad (2)$$

$$\sum_{i=1}^N x_i = 1 \quad (3)$$

$$x_i \geq 0 \quad (4)$$

Рассмотрим случай, когда критериальное значение доходности r^* удовлетворяет условиям (5)

$$\sum_{i=1}^N x_i r_{i1} \leq r^* \leq \sum_{i=1}^N x_i \tilde{r}_i = \tilde{r} \quad (5)$$

Тогда величина риска равна

$$\beta(x) = \frac{1}{\sum_{i=1}^N x_i r_{i2} - \sum_{i=1}^N x_i r_{i1}} \left[\left(r^* - \sum_{i=1}^N x_i r_{i1} \right) + \left(\sum_{i=1}^N x_i \tilde{r}_i - r^* \right) \right] \ln \left(\frac{\sum_{i=1}^N x_i \tilde{r}_i - r^*}{\sum_{i=1}^N x_i \tilde{r}_i - \sum_{i=1}^N x_i r_{i1}} \right) \quad (6)$$

Рассмотрим двойственную задачу оптимизации нечеткого портфеля относительно задачи (1)-(4) [4]:
минимизировать

$$\beta(x) \quad (7)$$

при условиях

$$\tilde{r} = \sum_{i=1}^N x_i \tilde{r}_i \geq r_{\text{зад}} = r^* \quad (8)$$

$$\sum_{i=1}^N x_i = 1, x_i \geq 0 \quad (9)$$

Рассмотрим теперь многокритериальную задачу, в которой будем минимизировать риск и максимизировать доходность портфеля.

Для того, чтобы определить структуру соответствующего портфеля нужно решить следующую задачу:

$$\{x_{\text{opt}}\} = \{x\} \mid r \rightarrow \max, \beta \rightarrow \min \quad (10)$$

где r и β определяются из формул (8) - (9),

а x удовлетворяет условию $\sum_{i=1}^N x_i = 1, x_i \geq 0$.

Доходность портфеля:

$$r = (r_{\min} = \sum_{i=1}^N x_i r_{1i}; \tilde{r} = \sum_{i=1}^N x_i \tilde{r}_i; r_{\max} = \sum_{i=1}^N x_i r_{2i}),$$

где $(r_{1i}, \tilde{r}_i, r_{2i})$ - доходность i -ой ценной бумаги- нечеткое число с известной функцией принадлежности.

Для решения данной используем способ приведения ее к однокритериальной. Для этого пронормируем значение доходности следующим образом:

$$\tilde{r}_n = \frac{r_{\max} - \tilde{r}}{r_{\max} - r_{\min}}, \quad \tilde{r}_n \in [0; 1]$$

Таким образом, получим оптимизационную задачу в следующем виде:

$$\begin{aligned} \{w_1 \tilde{r}_n + w_2 \beta(x)\} &\rightarrow \min \\ w_1 \geq 0, \quad w_2 \geq 0, \quad w_1 \neq w_2, \quad w_1 + w_2 &= 1 \\ \sum_{i=1}^N x_i &= 1 \quad x_i \geq 0 \quad i = \overline{1, N}. \end{aligned} \tag{11}$$

Рассматриваем. $0 < \beta < 1$. Это возможно в двух случаях:

когда $\sum_{i=1}^N x_i r_{i1} \leq r^* < \sum_{i=1}^N x_i \tilde{r}_i$, либо когда $\sum_{i=1}^N x_i \tilde{r}_i \leq r^* < \sum_{i=1}^N x_i r_{i2}$.

а) Пусть $\sum_{i=1}^N x_i r_{i1} \leq r^* < \sum_{i=1}^N x_i \tilde{r}_i$. Используя (8 - (9) задача сводится к следующей:

при условиях $\{w_1 \tilde{r}_n + w_2 \beta(x)\} \rightarrow \min$
 $w_1 \geq 0, \quad w_2 \geq 0, \quad w_1 \neq w_2, \quad w_1 + w_2 = 1$

$$\frac{1}{\sum_{i=1}^N x_i r_{i2} - \sum_{i=1}^N x_i r_{i1}} \left(\left(r^* - \sum_{i=1}^N x_i r_{i1} \right) + \left(\sum_{i=1}^N x_i \tilde{r}_i - r^* \right) \cdot \ln \left(\frac{\sum_{i=1}^N x_i \tilde{r}_i - r^*}{\sum_{i=1}^N x_i \tilde{r}_i - \sum_{i=1}^N x_i r_{i1}} \right) \right) = \beta(x),$$

$$\sum_{i=1}^N x_i r_{i1} \leq r^*, \quad \sum_{i=1}^N x_i \tilde{r}_i > r^*,$$

$$\sum_{i=1}^N x_i = 1 \quad x_i \geq 0 \quad i = \overline{1, N}.$$

б) Пусть $\sum_{i=1}^N x_i \tilde{r}_i \leq r^* < \sum_{i=1}^N x_i r_{i2}$, тогда задача сводится к следующей:

при условиях $\{w_1 \tilde{r}_n + w_2 \beta(x)\} \rightarrow \min$
 $w_1 \geq 0, \quad w_2 \geq 0, \quad w_1 \neq w_2, \quad w_1 + w_2 = 1$

$$\frac{1}{\sum_{i=1}^N x_i r_{i2} - \sum_{i=1}^N x_i r_{i1}} \left(\left(r^* - \sum_{i=1}^N x_i r_{i1} \right) - \left(r^* - \sum_{i=1}^N x_i \tilde{r}_i \right) \cdot \ln \left(\frac{r^* - \sum_{i=1}^N x_i \tilde{r}_i}{\sum_{i=1}^N x_i r_{i2} - \sum_{i=1}^N x_i r_{i1}} \right) \right) = \beta,$$

$$\sum_{i=1}^N x_i r_{i2} > r^*, \quad \sum_{i=1}^N x_i \tilde{r}_i \leq r^*,$$

$$\sum_{i=1}^N x_i = 1 \quad x_i \geq 0 \quad i = \overline{1, N}.$$

Экспериментальные исследования

Для проведения анализа исследуемого метода оптимизации инвестиционного портфеля был проанализирован фондовый рынок в периоды экономической стабильности и кризиса. Данные использованы из архива Московской фондовой биржи (МФБ) по торговле ценными бумагами компаний ОАО «Полет» (POLT), ОАО «Новые горизонты» (NGOR), ОАО «КОМСТАР-ОТС» (CMST), ОАО «Ситроникс» (SITR)].

Рассмотрим портфель из 4 активов. Используем данные по доходности акций за период с 02.04.07 по 03.08.07. Исходя из состояния фондового рынка делаем выводы о прибыльности акций:

- доходность акций POLT лежит в расчетном коридоре [0; 4,4], наиболее ожидаемое значение доходности 1,2%;
- доходность акций NGOR лежит в расчетном коридоре [-4,6; 1,13], наиболее ожидаемое значение доходности 0%;
- доходность акций CMST лежит в расчетном коридоре [-1,04; 1,02], наиболее ожидаемое значение доходности 0%;
- доходность акций SITR лежит в расчетном коридоре [-0,74; 1,96], наиболее ожидаемое значение доходности 0%;

Задав уровень критической прибыльности, и изменяя значения весовых коэффициентов, были получены следующие результаты, представленные в таблицах 1-4.

Таблица 1. Оптимальный портфель с критическим уровнем доходности 0,5%

POLT	NGOR	CMST	SITR	Ожидаемое значение доходности	Нижняя граница	Верхняя граница	Риск	w_2
0,91992	0,02501	0,02739	0,02768	1,1039	-0,16403	4,15809	0,04902	0,1
0,91948	0,02535	0,02746	0,02771	1,10338	-0,16565	4,15669	0,0492	0,2
0,91905	0,02567	0,02753	0,02775	1,10286	-0,16727	4,15529	0,04938	0,3
0,9186	0,02601	0,0276	0,02779	1,10233	-0,1689	4,15389	0,04956	0,4

Таблица 2. Оптимальный портфель с критическим уровнем доходности 1,5%

POLT	NGOR	CMST	SITR	Ожидаемое значение доходности	Нижняя граница	Верхняя граница	Риск	w_2
0,92352	0,02365	0,02599	0,02684	1,10822	-0,15569	4,16931	0,59292	0,1
0,92335	0,02391	0,02599	0,02675	1,10801	-0,15683	4,16868	0,59361	0,2
0,92317	0,02418	0,026	0,02665	1,10781	-0,15797	4,16804	0,5943	0,3
0,923	0,02444	0,026	0,02656	1,10761	-0,15911	4,16741	0,59499	0,4

Проведем теперь эксперименты для двухкомпонентного портфеля, задав критериальное значение прибыльности и изменяя значение весового коэффициента для второго критерия w_2 . Соответствующие результаты приведены в таблицах 3,4.

Таблица 3. Оптимальный портфель с критическим уровнем доходности 0,5%

POLT	NGOR	Ожидаемое значение доходности	Нижняя граница	Верхняя граница	Риск	W_2
0,94688	0,05312	1,13626	-0,24435	4,22630	0,05524	0,1
0,94667	0,05333	1,13601	-0,24531	4,22562	0,05534	0,2
0,94646	0,05354	1,13576	-0,24627	4,22494	0,05545	0,3
0,94625	0,05375	1,13551	-0,24723	4,22425	0,05555	0,4
0,94605	0,05395	1,13525	-0,24819	4,22357	0,05566	0,5

Таблица 4. Оптимальный портфель с критическим уровнем доходности 2%

POLT	NGOR	Ожидаемое значение доходности	Нижняя граница	Верхняя граница	Риск	W_2
0,94791	0,05209	1,13749	-0,23963	4,22966	0,81203	0,1
0,94781	0,05219	1,13737	-0,24009	4,22932	0,81234	0,2
0,94669	0,05331	1,13603	-0,24523	4,22567	0,81239	0,3
0,94659	0,05341	1,13591	-0,24569	4,22535	0,81244	0,4
0,94649	0,05351	1,13579	-0,24614	4,22503	0,81274	0,5

Таким образом, был построен оптимальный портфель с минимальным уровнем риска и максимальной доходностью (при $W_2 = 0,1$). Аналогичные результаты наблюдаем при построении портфеля из двух компонент.

При этом с увеличением уровня критической доходности соответственно увеличивается значение риска портфеля, так для портфеля из четырех компонент при $r^* = 0,5$, $\beta = 4,9\%$ а при $r^* = 1,5$, $\beta = 59\%$, для портфеля из двух компонент при $r^* = 0,5$, $\beta = 5,5\%$, а при $r^* = 2$, $\beta = 81\%$. Как и в предыдущих работах, посвященных данной тематике, зависимость доходность-риск приобретает ниспадающий характер, чем больше риск - тем меньше доходность, в отличие от вероятностных методов. Это объясняется тем, что в нечетко-множественном методе под риском понимается ситуация, когда ожидаемая доходность портфеля окажется ниже заданного критического уровня, со снижением ожидаемой доходности увеличивается риск того, что прибыль от портфельных инвестиций окажется ниже критического значения.

Теперь рассмотрим данные в период экономического кризиса с 27.10.08 по 27.02.09. Исходя из состояния фондового рынка делаем выводы о прибыльности акций:

- доходность акций POLT лежит в расчетном коридоре $[-0,18873; 0,4823]$, наиболее ожидаемое значение доходности 0,02%; - доходность акций NGOR лежит в расчетном коридоре $[-0,1686; 0,42053]$, наиболее ожидаемое значение доходности 0,07%; - доходность акций CMST лежит в расчетном коридоре $[-14,5059; 0,00054]$, наиболее ожидаемое значение доходности 0%; - доходность акций SITR лежит в расчетном коридоре $[-20,9484; 0]$, наиболее ожидаемое значение доходности -1%.

Задавая уровень критической прибыльности и регулируя весовые коэффициенты, получим результаты, представленные в таблице 5 и таблице 6 (для четырех компонент).

Таблица 5. Оптимальный портфель с критическим уровнем доходности -0,5%

POLT	NGOR	CMST	SITR	Ожидаемое значение доходности	Нижняя граница	Верхняя граница	Риск	w_2
0,0352	0,92893	0,02162	0,01425	0,05148	-0,77536	0,40763	0,0441	0,1
0,03419	0,92857	0,02195	0,01529	0,05039	-0,80174	0,40699	0,05071	0,2
0,03417	0,92444	0,02359	0,0178	0,04759	-0,87742	0,40525	0,07057	0,3
0,03314	0,92394	0,02397	0,01895	0,04639	-0,90662	0,40454	0,07842	0,4
0,03213	0,92333	0,02439	0,02015	0,04512	-0,93775	0,40379	0,08687	0,5
0,0311	0,92263	0,02485	0,02142	0,04379	-0,97066	0,40301	0,09586	0,6
0,03008	0,92183	0,02535	0,02274	0,04239	-1,00522	0,40218	0,10532	0,7

Таблица 6. Оптимальный портфель с критическим уровнем доходности 0%

POLT	NGOR	CMST	SITR	Ожидаемое значение доходности	Нижняя граница	Верхняя граница	Риск	w_2
0,03397	0,9274	0,02549	0,01314	0,05245	-0,80784	0,4064	0,54281	0,1
0,03288	0,92774	0,02526	0,01412	0,05148	-0,82492	0,40601	0,55029	0,2
0,03277	0,92417	0,02636	0,0167	0,04865	-0,89424	0,40446	0,57534	0,3
0,03167	0,92447	0,02614	0,01772	0,04762	-0,91221	0,40406	0,58252	0,4
0,03057	0,9247	0,02592	0,01881	0,04653	-0,93169	0,40362	0,59009	0,5
0,03045	0,92146	0,02687	0,02122	0,04389	-0,99539	0,4022	0,61053	0,6
0,02935	0,92167	0,02664	0,02234	0,04276	-1,01539	0,40176	0,61771	0,7

Для двух компонент

Таблица 7. Оптимальный портфель с критическим уровнем доходности -0,5%

POLT	NGOR	Ожидаемое значение доходности	Нижняя граница	Верхняя граница	Риск	w_2
0,95129	0,04871	0,01903	-0,88609	0,45883	0,07267	0,1
0,95121	0,04879	0,01902	-0,88728	0,45879	0,07297	0,2
0,95105	0,04895	0,01902	-0,8895	0,45872	0,07354	0,3
0,9496	0,0504	0,01899	-0,91034	0,45802	0,07911	0,4
0,94934	0,05066	0,01899	-0,91405	0,45789	0,08008	0,5
0,94901	0,05099	0,01898	-0,91877	0,45773	0,08133	0,6
0,94749	0,05251	0,01895	-0,94045	0,45701	0,08716	0,7

Таблица 8. Оптимальный портфель с критическим уровнем доходности 0%

POLT	NGOR	Ожидаемое значение доходности	Нижняя граница	Верхняя граница	Риск	w_2
0,95061	0,04939	0,02301	-0,89583	0,45851	0,6052	0,1
0,95078	0,04922	0,02302	-0,89348	0,45859	0,60476	0,2
0,9496	0,0504	0,02249	-0,91028	0,45802	0,60942	0,3
0,94968	0,05032	0,02239	-0,9091	0,45806	0,60931	0,4
0,94852	0,05148	0,02227	-0,92584	0,4575	0,61385	0,5
0,94737	0,05263	0,02215	-0,94225	0,45694	0,6182	0,6
0,94736	0,05264	0,02212	-0,94243	0,45694	0,61846	0,7

Анализируя результаты в табл. 5-8, мы видим, что ожидаемое значение доходности портфеля, заметно снизилось по сравнению с предыдущим случаем-периодом экономической стабильности. Кроме того, как и прежде зависимость «оптимальная доходность-риск» является монотонно убывающей. С ростом критериального значения доходности риск возрастает, что хорошо согласуется с теоретическими положениями.

Заключение

В работе рассмотрена и исследована многокритериальная задача нечеткой портфельной оптимизации.

Предложен метод сведения ее к однокритериальной задаче портфельной оптимизации. Проведены экспериментальные исследования полученной модели на примере рынка российских компаний и выполнен анализ полученных решений.

Определены зависимости ожидаемой доходности нечеткого портфеля от величины риска, а также зависимости риска от критериального значения доходности. Предлагаемый подход позволяет найти наилучший компромисс между ожидаемой доходностью портфеля и его риском в условиях неопределенности.

Литература

- [Зайченко, 2007] Зайченко Юрий, Малихех Есфандиярфард. Анализ и сравнение результатов оптимизации инвестиционного портфеля при применении модели Марковитца и нечетко-множественного метода. //Proceedings of X111-th International Conference KDS-2007 "Knowledge, Dialogue Solution", Vol.1, pp.278-2.
- [Зайченко, 2008] Зайченко Ю.П., Малихех Есфандиярфард. Оптимизация инвестиционного портфеля в условиях неопределенности// Системні дослідження та інформаційні технології.-№2-2008.-с. 59-76.
- [Зайченко, Малихех, 2007] Зайченко Ю.П., д.т.н., проф.,Малихех Есфандиярфард,Заика А.И. Анализ инвестиционного портфеля на основе прогнозирования курсов акций // Вісник національного технічного університету України «КПІ». Інформатика, управління та обчислювальна техніка.» Київ ТОО «ВЕК+», №47 – 2007, - С. 168-179.
- [Зайченко, 2010] Зайченко Юрий, Ови Нафас Агаи Аг Гамиш. Исследование двойственной задачи оптимизации инвестиционного портфеля в нечетких условиях. Natural and Artificial Intelligence. ITHEA. Sofia, Bulgaria.-2010.-pp. 115-128.

Информация об авторах

Юрий Зайченко– доктор технических наук, профессор. Институт прикладного системного анализа НТУУ «КПИ», 03056, Киев-56, Украина phone: 38044 -4068393, e-mail: baskervil@voliacable.com, ZaychenkoYuri@ukr.net

Esfandiaryfard Maliheh - IRAN -Semnan Electricity Company phone 09191744969, e-mail: far_d_sem@yahoo.com

Ови Нафас Агаи Аг Гамиш (Иран) - аспирант НТУУ «КПИ»; 03056, Киев-56, Украина e-mail: ovinafas@yahoo.com

Благодарности

Статья частично финансирована из проекта ITHEA XXI Института Информационных теорий и Приложений FOI ITHEA и консорциума FOI Bulgaria (www.ithea.org, www.foibg.com)

О СВОЙСТВАХ РАССТОЯНИЙ В ВЕРОЯТНОСТНЫХ ВЫСКАЗЫВАНИЯХ ЭКСПЕРТОВ *

Александр Викентьев, Руслан Викентьев

Abstract: The paper discusses probabilities' logical expert statements represented as the formulas of Sentence Logic. Methods for setting metrics on such formulas are offered and the entered metric's properties are investigated. The research can be applied to solving the problems of the best reconciliation of expert statements, to constructing the decision functions in pattern recognition and building the expert systems.

Keywords: *cluster analysis, expert statements, distance, metrics*

ACM Classification Keywords: *I.2.6. Artificial Intelligence - knowledge acquisition.*

Введение

Поскольку сейчас проявляется все больший интерес к анализу экспертной информации, заданной в виде вероятностных логических высказываний экспертов, интересны также вопросы о высказываниях экспертов, представленных формулами исчисления высказываний с вероятностями. Возникают задачи об алгоритмах распознавания закономерностей, согласования таких знаний и их кластеризации [Блощицын В.Я., Лбов, 1987 - Викентьев, Лбов, 1998]. Для этого необходимы метрики на знаниях. В работе рассматриваются логические высказывания экспертов, представленные формулами исчисления высказываний (ИВ) с вероятностями.

Вопросами введения расстояний на высказываниях экспертов и их применениями занимался профессор Г.С. Лбов [Лбов, Старцева, 1999], использовал для построения решающих функций и для согласования высказываний совместно со своими учениками, использует в своих работах д.т.н. В.Б. Бериков [Бериков, 2009 -- Лбов, Бериков, 2005], разрабатывал и продолжает применять в алгоритмах распознавания профессор Н.Г. Загоруйко и один из авторов статьи.

В данной статье предлагаются способы задания расстояний на логических высказываниях с вероятностями с использованием логических моделей, вообще говоря, неполной теории. Если класс моделей теории эффективно задан, то мы можем все нужные расстояния найти, а также вычислить вероятности выполнимости любой формулы на данном конечном классе. Нами здесь рассматривается обратная задача: мы знаем от экспертов вероятности формул-высказываний, знаем как-то логическую теорию, которая задает модели нашей исследуемой области и хотим восстановить расстояния между данными формулами, мы его не знаем и хотим узнать. Ясно, что такая задача может иметь несколько решений из-за нехватки информации и про теорию и про адекватность ее моделей. В случае когда теория одна и модели ее устраивают всех экспертов, то решение будет единственно.

Изучаются свойства введенных искомым расстояний. Для решения поставленной задачи используются вероятностный и теоретико - модельный подходы [Лбов, Старцева, 1999 -- Vikent'ev, Lbov, 1997]. Результаты неоднократно доложены на международных конференциях в 2009, 2010 гг., в том числе юбилейной конференции к 100 - летию Академика Анатолия Ивановича Мальцева – выдающегося математика -- Логика и Алгебраиста, крупного ученого по алгебре, теории моделей, теории алгоритмов, теории нумерации, алгоритмическим проблемам и применениям методов математической логики.

* Работа выполнена при поддержке гранта РФФИ 10-01-00113а

РАССТОЯНИЯ НА ФОРМУЛАХ ИВ С ВЕРОЯТНОСТЯМИ И ИХ СВОЙСТВА

Будем рассматривать знания экспертов, представленные формулами ИВ с вероятностями (вероятностные высказывания), т.е. высказывания вида: " ϕ с вероятностью p_ϕ ", где ϕ - формула ИВ.

Используем запись для таких высказываний: $B_i = \langle \phi, p_\phi \rangle$, $B_j = \langle \psi, p_\psi \rangle$.

Знакомство с математической логикой предполагается [Ершов, Палютин, 2004]

Пусть Σ - база знаний, состоящая из формул ИВ (в Σ содержатся все формулы, с которыми будут работать эксперты, а также конечное число аксиом теории). $S(\phi)$ - носитель формулы ϕ , т.е. множество элементарных высказываний, используемых при написании формулы ϕ . $S(\Sigma) = \bigcup_{\phi \in \Sigma} S(\phi)$ - носитель

совокупности знаний. Для простоты будем считать, что он конечен.

Рассмотрим множество $P(S(\Sigma)) = 2^{S(\Sigma)}$ - множество всевозможных подмножеств множества $S(\Sigma)$. Элементы множества $P(S(\Sigma))$ назовем моделями. Известно, что

$|P(S(\Sigma))| = 2^{|S(\Sigma)|} = n$ (-- для простоты обозначения). Модели теории образуют подмножество этого множества и для его мощности, в дальнейшем, будем использовать тоже n .

Пусть эксперты говорят о вероятностях (частоты) формул на множестве n моделей, и каждое высказывание присутствует только с одной вероятностью.

Тогда будем интерпретировать вероятность, данную экспертом, следующим образом: $B = \langle \phi, p_\phi \rangle$

означает, что высказывание ϕ истинно на $n_\phi = \lfloor n \cdot p_\phi \rfloor$ моделях, где n - число моделей рассматриваемой теории..

Пусть у нас есть два вероятностных логических высказывания $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$.

Зададим способ вычисления расстояния $\rho(B_i, B_j)$ между такими высказываниями.

Интерпретируя данные экспертами вероятности описанным выше способом, получаем, что высказывание ϕ истинно на $n_\phi = \lfloor n \cdot p_\phi \rfloor$ моделях, высказывание ψ истинно на $n_\psi = \lfloor n \cdot p_\psi \rfloor$ моделях.

Предполагаем, что с этим подходом (аксиомами теории и интерпретацией вероятностей) согласны сами эксперты.

Отметим также, что (при таком подходе) мы точно не знаем на каких именно моделях каждое высказывание истинно, а также не знаем число моделей, на которых эти высказывания-пары истинны одновременно.

Будем решать такую задачу: пусть высказывание ϕ истинно на n_ϕ моделях, высказывание ψ истинно на n_ψ моделях и k - число моделей, на которых эти высказывания истинны одновременно. Тогда как

находить и вычислять расстояние между высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$?

Обозначим возникающие далее расстояния за $\rho_k(B_i, B_j)$, где $k = 0, 1, 2, \dots, \min(n_\phi, n_\psi)$.

Как и раньше [Лбов, Старцева, 1999 -- Викентьев, Лбов, 1998], расстояние $\rho_k(B_i, B_j)$ определим через симметрическую разность моделей на которых они истинны, то есть

$$\rho_k(B_i, B_j) = \frac{n_\phi - k + n_\psi - k}{n} = \frac{n_\phi + n_\psi - 2k}{n}, \quad \text{для каждого}$$

$$k = 0, 1, 2, \dots, \min(n_\phi, n_\psi).$$

Теорема 1. Для расстояний $\rho_k(B_i, B_j)$ справедливы свойства:

$$0 \leq \rho_k(B_i, B_j) \leq 1.$$

$$\rho_k(B_i, B_j) = \rho_k(B_j, B_i).$$

$$\rho_k(B_i, B_j) \leq \rho_k(B_i, B_s) + \rho_k(B_s, B_j).$$

$B_i \equiv B_j \Leftrightarrow \rho_k(B_i, B_j) = 0$ ($B_i \equiv B_j \Leftrightarrow \phi \equiv \psi$ и $p_\phi = p_\psi$ - это означает, что формулы ϕ и ψ истинны на одних и тех же моделях).

$$B_i \equiv \neg B_j \Leftrightarrow \rho_k(B_i, B_j) = 1.$$

$$\rho_k(B_i, B_j) = 1 - \rho_k(B_i, \neg B_j) = \rho_k(\neg B_i, \neg B_j).$$

$$\rho_k(B_i, B_j) = \rho_k(B_i \wedge B_j, B_i \vee B_j).$$

Докажем неочевидное свойство 3. Определение $\rho_k(B_i, B_j)$ можно переписать следующим образом:

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n} = \frac{n_\phi \Delta \psi}{n} = \frac{n_{(\neg \phi \wedge \psi) \vee (\phi \wedge \neg \psi)}}{n} \quad (\text{по определению}$$

симметрической разности). Тогда для произвольного высказывания $B_s = \langle \chi, p_\chi \rangle$ нетрудно доказываем, что $n_\phi \Delta \psi \leq n_\phi \Delta \chi + n_\chi \Delta \psi$. Тогда

$$\rho_k(B_i, B_j) = \frac{n_\phi \Delta \psi}{n} \leq \frac{n_\phi \Delta \chi}{n} + \frac{n_\chi \Delta \psi}{n} = \rho_k(B_i, B_s) + \rho_k(B_s, B_j).$$

Докажем так же свойство 4. Докажем сначала слева направо (\Rightarrow). Если $B_i \equiv B_j$, то $\phi \equiv \psi$ и,

$$\text{значит, } n_\phi = n_\psi = k. \text{ Следовательно, } \rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n} = 0.$$

Докажем теперь в обратную сторону (\Leftarrow). Если $\rho_k(B_i, B_j) = 0$, то $n_\phi + n_\psi - 2k = 0$. Так как k может принимать значения $0, 1, 2, \dots, \min(n_\phi, n_\psi)$, то $(n_\phi + n_\psi - 2k = 0 \Leftrightarrow n_\phi = n_\psi = k)$. Следовательно, $\phi \equiv \psi$ и значит $B_i \equiv B_j$. Докажем

свойство 5. Докажем (\Rightarrow). Если $B_i \equiv \neg B_j$, то $\phi \equiv \neg \psi$. Тогда $n_\phi = n - n_\psi$ и $k = 0$,

$$\text{следовательно, } \rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 0}{n} = \frac{n}{n} = 1.$$

Докажем (\Leftarrow). Если $\rho_k(B_i, B_j) = 1$, то $n_\phi + n_\psi - 2k = n$. Так как k может принимать значения $0, 1, 2, \dots, \min(n_\phi, n_\psi)$, то $(n_\phi + n_\psi - 2k = n \Leftrightarrow n_\phi + n_\psi = n$ и $k = 0)$.

Следовательно, $\phi \equiv \neg \psi$ и $B_i \equiv \neg B_j$.

Докажем свойство 6. $\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n}$. Рассмотрим

$$\begin{aligned} \rho_k(B_i, \neg B_j) &= \frac{n_\phi + n_{\neg\psi} - 2(n_\phi - k)}{n} = \frac{n_\phi + (n - n_\psi) - 2(n_\phi - k)}{n} = \\ &= \frac{n - n_\phi - n_\psi - 2k}{n} = 1 - \frac{n_\phi + n_\psi - 2k}{n} = 1 - \rho_k(B_i, B_j). \end{aligned}$$

Доказывается нетрудно, что $1 - \rho_k(B_i, \neg B_j) = \rho_k(\neg B_i, \neg B_j)$.

Докажем свойство 7. Можно доказать, что $n_{(\phi \wedge \psi) \Delta (\phi \vee \psi)} = n_{\phi \Delta \psi}$. Тогда

$$\rho_k(B_i \wedge B_j, B_i \vee B_j) = \frac{n_{(\phi \wedge \psi) \Delta (\phi \vee \psi)}}{n} = \frac{n_{\phi \Delta \psi}}{n} = \rho_k(B_i, B_j).$$

Далее предложим несколько способов вычисления расстояния $\rho(B_i, B_j)$ между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$.

Так как нам не известно число k (число моделей, на которых высказывания ϕ и ψ истинны одновременно), и если нет никаких предпочтений для значения k (хотя оно и может быть высказано экспертами), то можем, например, поступить следующим образом.

Предположим так же, что для нас все значения для числа k равновероятны. Тогда расстояние между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$ определим как усреднение

$$\text{расстояний } \rho_k(B_i, B_j) \text{ по всем значениям } k, \text{ т.е. } \rho(B_i, B_j) = \frac{\sum_{k=0}^{\min(n_\phi, n_\psi)} \rho_k(B_i, B_j)}{\min(n_\phi, n_\psi) + 1}.$$

Для этого расстояния также справедлива теорема 1, и под знаком суммы слагаемые можно взять с весами с учетом общих мнений экспертов или адаптацией по задаче.

Если экспертами высказано, какое значение для k предпочтительнее, то в качестве $\rho(B_i, B_j)$ берем расстояние $\rho_k(B_i, B_j)$. Это так когда мы знаем, что пересечение состоит из k моделей.

Можно подойти к этому вопросу и с вероятностной точки зрения, построить статистико-вероятностную модель для каждого k и вычислить вероятность того, что высказывания ϕ и ψ одновременно истинны на k моделях.

Далее найдем вероятность p_k того, что в выбранных n_ϕ моделях и n_ψ моделях (они выбираются из n моделей) будет k моделей, на которых высказывания ϕ и ψ истинны одновременно, где $k = 0, 1, 2, \dots, \min(n_\phi, n_\psi)$.

Сначала определим вероятностное пространство $\langle \Omega, \mathcal{A}, p \rangle$, где Ω - пространство элементарных исходов - моделей, $\Omega = \{0, 1, 2, \dots, \min(n_\phi, n_\psi)\}$ - число возможных совпадений моделей в наборах из n_ϕ и n_ψ моделей, \mathcal{A} - система пар подмножеств множества моделей Ω , образующая σ - алгебру событий, и p - вероятность на (Ω, \mathcal{A}) .

Определим на $\Omega \times A$ случайную величину ξ так: $\xi(k) = \rho_k(B_i, B_j)$, $k = 0, 1, 2, \dots, \min(n_\phi, n_\psi)$, т.е. ξ каждому k из Ω ставит в соответствие расстояние $\rho_k(B_i, B_j)$.

Вероятность появления этого события- расстояния для k на классе из n моделей можно вычислить так:

$$p_k = \frac{C_n^k C_{n-k}^{n_\phi-k} C_{n-n_\phi}^{n_\psi-k}}{C_n^{n_\phi} C_n^{n_\psi}}, \text{ где } C_n^{n_\phi} - \text{число способов выбрать } n_\phi \text{ моделей из } n \text{ моделей.}$$

Действительно, так как любое множество (набор моделей), состоящее из n_ϕ моделей, может сочетаться с любым набором моделей, состоящим из n_ψ моделей, то число $(C_n^{n_\phi} \cdot C_n^{n_\psi})$ - количество всех способов выбрать два набора моделей, один из которых состоит из n_ϕ моделей, а другой из n_ψ моделей.

Выбрать k моделей, которые будут общими в этих наборах, из n моделей можно C_n^k способами. Тогда остальные $(n_\phi - k)$ и $(n_\psi - k)$ моделей в наборах должны быть дизъюнктивными. Следовательно, остальные $(n_\phi - k)$ моделей для пополнения

набора, состоящего из k до n_ϕ моделей, можно выбрать $C_{n-k}^{n_\phi-k}$ способами, а $(n_\psi - k)$ моделей для получения набора, состоящего из n_ψ моделей, с учетом наших предположений, -- $C_{n-n_\phi}^{n_\psi-k}$ способами.

Значит, имеется всего $(C_n^k C_{n-k}^{n_\phi-k} C_{n-n_\phi}^{n_\psi-k})$ способов выбрать два набора, один из которых состоит из n_ϕ моделей, а другой из n_ψ , и в точности k моделей в этих наборах моделей совпадают. Поэтому вероятность того, что k моделей совпадет в наборах из n_ϕ и n_ψ элементарных моделей, будет

$$\text{равна } p_k = \frac{C_n^k C_{n-k}^{n_\phi-k} C_{n-n_\phi}^{n_\psi-k}}{C_n^{n_\phi} C_n^{n_\psi}}.$$

В результате получим, что расстояния $\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n}$ будут

$$\text{появляться с вероятностями } p_k = \frac{C_n^k C_{n-k}^{n_\phi-k} C_{n-n_\phi}^{n_\psi-k}}{C_n^{n_\phi} C_n^{n_\psi}}, \text{ где } k = 0, 1, 2, \dots, \min(n_\phi, n_\psi).$$

Предлагается эти вероятности или ближайшие к ним числа использовать в качестве весов расстояний для получения результирующего расстояния для данных вероятностных формул. Более подробно об этом будет немного дальше. Заметим, что при таком подходе главную роль играют не сами формулы, а числа определяющие количество моделей и их пересечения. Не имея другой информации, мы посчитали все

интересующие нас подмножества для вычисления частоты (или вероятности) появления расстояния для конкретного k . Используя свойство инвариантности расстояний между формулами и вероятностей высказываний (формул) [Лбов, Старцева,1999 -- Викентьев, Лбов, 1998], можно только что проведенные рассуждения проделать с меньшим носителем знаний: включающем только те модели теории, которые включают носители встречающиеся в двух формулах, для которых ищется расстояние. Будем считать, что мы так сделали с самого начала. И тогда подсчет в этом случае будет оптимальным.

Зная вероятности p_k для каждого расстояния $\rho_k(B_i, B_j)$, в качестве расстояния между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$ можно взять, например, наиболее вероятное $\rho(B_i, B_j) = \rho_m(B_i, B_j)$, где $p_m = \max_k p_k$. Для такого $\rho(B_i, B_j)$

справедлива теорема 1. Для получения другого расстояния можно взять усредненное некоторого подмножества (отобранных по предпочтению) из полученных.

Более общо, беря произвольные p_k (исходя из экспертных оценок или дополнительных сведений экспертов, которые могут и не совпадать) для расстояний $\rho_k(B_i, B_j)$ так, чтобы получался закон распределения, получим самый общий случай для адаптивного поиска нужного расстояния между формулами с вероятностями.

Тогда в качестве расстояния между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$ можно взять величину, равную математическому ожиданию (центру тяжести) или среднему значению случайной величины ξ , т.е.

$$\rho(B_i, B_j) = M\xi = \sum_{k=0}^{\min(n_\phi, n_\psi)} \rho_k(B_i, B_j) \cdot p_k.$$

Для так введенного расстояния справедлива

следующая

Теорема 2. Для расстояния $\rho(B_i, B_j)$ справедливы свойства:

$$0 \leq \rho(B_i, B_j) \leq 1.$$

$$\rho(B_i, B_j) = \rho(B_j, B_i).$$

$$\rho(B_i, B_j) \leq \rho(B_i, B_s) + \rho(B_s, B_j).$$

Если $\rho(B_i, B_j) = 0$, то $B_i \equiv B_j$.

$$\rho(B_i, B_j) = 1 - \rho(B_i, \neg B_j) = \rho(\neg B_i, \neg B_j).$$

$$\rho(B_i, B_j) = \rho(B_i \wedge B_j, B_i \vee B_j).$$

Доказательство. Докажем сначала свойство 3. Для расстояний на $B_i = \langle \phi, p_\phi \rangle$ по свойству 3 теоремы 1 имеем $\rho_k(B_i, B_j) \leq \rho_k(B_i, B_s) + \rho_k(B_s, B_j)$. Тогда по свойствам математического ожидания {а) $\xi \leq \eta \Rightarrow M\xi \leq M\eta$, б) $M(\xi + \eta) = M\xi + M\eta$ и получаем требуемое свойство для расстояния.

Докажем свойство 4. Пусть $\xi(k) = \rho_k(B_i, B_j) \geq 0$ и $M\xi = 0$. Тогда по свойству математического ожидания $\xi(k) = \rho_k(B_i, B_j) = 0$ с вероятностью равной 1. Тогда по свойству 4

теоремы 1 $B_i \equiv B_j$. Докажем свойство 5. Так как для расстояния $\rho_k(B_i, B_j)$ по свойству 6 теоремы 1 справедливо равенство $\rho_k(B_i, B_j) = 1 - \rho_k(B_i, \neg B_j) = \rho_k(\neg B_i, \neg B_j)$, тогда по свойствам математического ожидания {а) если $p(\xi = \eta) = 1$ и $\exists M\xi$, то $M\xi = M\eta$, б) $M(a + b\xi) = a + bM\xi$ } получаем требуемое свойство для расстояния $\rho(B_i, B_j)$. Остальное доказывается аналогично.

Заключение

В работе предложены способы введения метрик на высказываниях экспертов -- формулах ИВ с вероятностями с помощью моделей логической теории исследуемой области. Исследование найдет применение в решении задач согласования вероятностных высказываний экспертов, кластеризации, и в построении баз знаний и экспертных систем. Результаты верны для формул над бесконечными носителями, переносятся на формулы с переменными языка 1-го порядка и формулы от разнотипных переменных с использованием измеримых подклассов (для фиксированной неполной теории) измеримых (в том числе и метрических) моделей.

Библиография

- [Блощицын В.Я., Лбов, 1987] Блощицын В.Я., Лбов Г.С. О мерах информативности логических высказываний. // Доклады Республиканской Школы-Семинара «Технология разработки экспертных систем». Кишинев, 1987, с.12-14.
- [Лбов, Старцева, 1999] Лбов Г.С., Старцева Н.Г. Логические решающие функции и вопросы статистической устойчивости решений. Новосибирск: Издательство Института математики, 1999, 212 с.
- [Vikent'ev, Lbov, 1997] Vikent'ev A.A., Lbov G.S.. Setting the metric and informativeness on statements of experts. // Pattern Recognition And Image Analysis. 1997, v. 7 (2), p. 175-189.
- [Викентьев, Лбов, 1998] Викентьев А.А., Лбов Г.С. О метризациях булевой алгебры предложений и информативности высказываний экспертов. // Доклады РАН, 1998, т. 361 (2), с.174-176
- [Бериков, 2009] Бериков В.Б. Кластерный анализ с использованием коллектива деревьев решений // Научный вестник НГТУ. 2009. № 3 (36). С.67-76.
- [Лбов, Бериков, 2005] Лбов Г.С., Бериков В.Б. Устойчивость решающих функций в задачах распознавания образов и анализа разнотипной информации. Новосибирск: Изд-во Ин-та математики, 2005. - 218 с.
- [Ершов, Палютин, 2004] Ершов Ю.Л., Палютин Е.А. Математическая логика. М.: Наука, 2004, 336 с.

Authors' Information

Александр А. Викентьев – с.н.с., канд. физ-мат. наук, Институт математики СО РАН, пр. Академика Коптюга, д.4, Лаборатория анализа данных. Доцент, Новосибирский госуд. университет; e-mail: vikent@math.nsc.ru

Руслан А. Викентьев – инженер, Институт математики СО РАН, пр. Академика Коптюга, д.4, Лаборатория анализа данных. Ассистент-преподаватель, Новосибирский госуд. университет; e-mail: ruslan.vikentiev@gmail.com

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ ПАРАМЕТРОВ РЕГИОНАЛЬНЫХ БЮДЖЕТОВ

Диана Омельянчик

Аннотация: Рассмотрена оптимизационная модель регионального бюджета и межбюджетных отношений. Предлагается математическая модель задачи оптимизации баланса величин налоговых ставок и ставок отчислений для региональных бюджетов, которая приводит к задачам непрерывной и комбинаторной оптимизации.

Ключевые слова: математическое моделирование, региональный бюджет, межбюджетные отношения, налоговые ставки, комбинаторная оптимизация.

ACM Classification Keywords: G.1.6: Mathematics of Computing: Numerical Analysis: Optimization; H.4.2 Information Systems Applications: Types of Systems: Decision Support.

Conference topic: Decision making. .

Введение

Одной из наиболее актуальных проблем как развивающихся стран, так и стран с переходной экономикой является недостаточная самостоятельность регионов в свете современных тенденций децентрализации государственного управления. Независимость регионального самоуправления невозможна без достаточной финансовой базы, способной удовлетворять потребности региона, а этого можно достигнуть только в результате эффективной бюджетной политики государства. Исходя из этого, создание математической модели регионального бюджета и межбюджетных отношений, с помощью которой можно адекватно решать проблемы оптимизации разнообразных финансово-экономических и социальных показателей, является необходимым условием успешности экономики всей страны.

Цель работы – описать одну из возможных математических моделей регионального бюджета и межбюджетных отношений, формализовать основные аспекты управления региональным бюджетом и критерии его оптимизации, сформулировать задачу поиска величины оптимальных налоговых ставок в виде задачи комбинаторной оптимизации.

Изучению механизмов бюджетного процесса и межбюджетных отношений посвящены научные исследования многих отечественных и зарубежных ученых. Построенные модели можно классифицировать с помощью двух характеристик: стадии бюджетного процесса и использованной математической концепции.

В основном, внимание исследователей привлекает первая стадия бюджетного процесса – прогнозирование и планирование. Разные подходы к построению подобных моделей представлены в работах [Sinnuany-Stern, 1993] и [Лукша, 1997]. Вторая (выполнение бюджета) и третья (анализ и мониторинг выполнения бюджета) стадии наиболее интересны с точки зрения создания разнообразных программно-алгоритмических комплексов (см. [Степине, 1986], [Нариньяни, 2008]) и введения специальных бюджетных коэффициентов, позволяющих анализировать эффективность государственного управления на региональном уровне. Также в эту классификацию стоит ввести задачи, посвященные расчету оптимального объема межбюджетного трансферта регионам, которые нуждаются в поддержке. Возможные подходы к ее решению можно найти в работах российских ученых [Соколова, 2002], [Россель, 1994], [Смирнов, 1996].

Спектр математических методов, которые применяются для выполнения заданий региональной бюджетной политики, очень широк. В ранних моделях бюджетов предлагается использование линейной регрессии, временных рядов, многокритериальной оптимизации и балансового подхода. С развитием таких смежных областей науки, как прикладная математика и математическая экономика, для описания бюджетных процессов начинают применять методы стохастической математики, теории нечетких множеств, теории игр, нейронные сети.

Исходя из предложенной классификации, в данной статье рассматривается модель планирования и прогнозирования регионального бюджета на основе методов комбинаторной оптимизации. Применение алгоритмов комбинаторной оптимизации к подобным задачам управления представляет интерес с научной и практической точки зрения.

Оптимизационная модель регионального бюджета и межбюджетных отношений

В данной статье предлагается следующая оптимизационная модель регионального бюджета и межбюджетных отношений (рис.1). Бюджет высшего уровня (государственный бюджет) рассматривается лишь в качестве звена, которое осуществляет перераспределение средств, поступивших из региональных бюджетов. Также предполагается, что бюджет региона сначала имеет только два источника доходов: налоги и доходы собственных предприятий. По результатам финансовой деятельности региона и в соответствии с поступлениями в бюджет высшего уровня, рассчитывается сумма трансферта соответствующему региональному бюджету. Расходы регионального бюджета состоят из отчислений в бюджет высшего уровня и затрат на финансирование бюджетных программ.

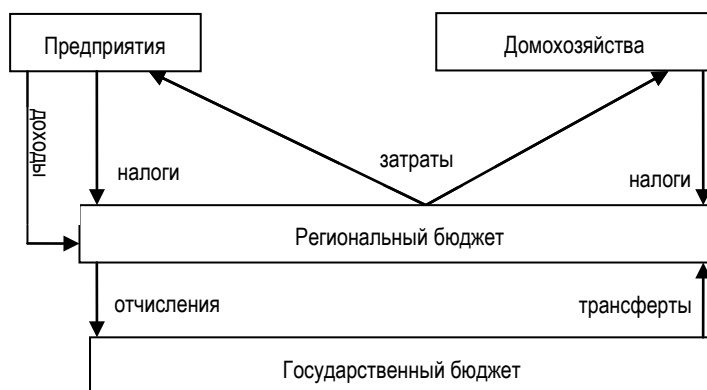


Рис. 1.

Для подобной упрощенной модели регионального бюджета можно поставить серию задач оптимизации. В качестве возможных параметров регулирования определяются: налоговая ставка; ставка отчислений в бюджет высшего уровня; доли статей расходов; минимально необходимая сумма расходов; корректирующие коэффициенты расчета трансферта. В зависимости от желаемых целей, критерии также могут быть разными. Например, минимизация дефицита бюджета или расчет оптимальной величины трансферта. Интересной задачей является выбор наилучшего сценария бюджетной политики, который можно осуществить с помощью регулирования целевых коэффициентов: показателя дотационности, уровня финансовой независимости региона, показателя донорства, показателя дефицитности бюджета, показателя бюджетного покрытия, показателя бюджетной задолженности, показателя бюджетной результативности территорий и т.д.

Перейдем к формальному описанию модели. Введем обозначения для экономических величин, участвующих в модели. Пусть:

n - количество статей налоговых поступлений в бюджет региона;

m - количество статей расходов бюджета;

S_i - сумма, подлежащая налогообложению по статье i ;

T_i - налоговая ставка i -ого налога;

g_i - ставка отчислений в бюджет высшего уровня по i -ой статье налогов;

v_j - доля расходов бюджета по статье j ;

m_j - минимальная сумма расходов по статье j , необходимая для функционирования региона;

L_{T_i}, U_{T_i} - нижняя и верхняя границы налоговой ставки по i -ому налогу;

L_{g_i}, U_{g_i} - нижняя и верхняя границы ставки отчислений по i -ой статье;

L_{v_j}, U_{v_j} - нижняя и верхняя границы доли расходов по статье j ;

P - доход собственных предприятий;

E - расчетная величина расходов бюджета;

C - корректирующий коэффициент;

D - величина межбюджетного трансферта;

H - среднегодовая численность населения региона.

Суммарные доходы региона I в рамках нашей модели состоят из налоговых поступлений и доходов собственных предприятий региона, что можно записать так:

$$I = \sum_{i=1}^n S_i T_i + P.$$

Суммарные расходы региона W состоят из расходов на реализацию бюджетных программ и отчислений налоговых поступлений в бюджет высшего уровня. В формальной постановке будем предполагать, что расходы на реализацию бюджетных программ задаются извне в виде расчетной величины расходов бюджета. Общее выражение для вычисления расходов W можно представить в таком виде:

$$W = \sum_{j=1}^m S_j T_j g_j + E.$$

Учитывая отчисления в бюджет высшего уровня, собственные доходы региона I' в рамках нашей модели можно записать так:

$$I' = \sum_{i=1}^n S_i T_i (1 - g_i) + P.$$

В случае, когда расчетные расходы бюджета превышают ожидаемые доходы ($I < W$), со стороны государственного бюджета региону предоставляется межбюджетный трансферт D :

$$D = C(W - I).$$

Доли статей расходов в итоговой сумме должны давать единицу:

$$\sum_{j=1}^m v_j = 1, v_j \geq 0, j = \overline{1, m}. \quad (1)$$

Расчетная величина расходов должна превышать суммарные минимальные расходы по бюджетным программам, необходимые для функционирования региона:

$$E = \sum_{j=1}^m v_j E \geq \sum_{j=1}^m m_j \quad (2)$$

Или

$$v_j E \geq m_j, j = \overline{1, m}. \quad (3)$$

Минимальные необходимые расходы также должны удовлетворять внутреннему условию непротиворечивости:

$$\frac{m_j}{\sum_{j=1}^m m_j} \geq v_j. \quad (4)$$

В завершение, добавим условия ограниченности налоговых ставок, ставок отчисления и долей статей расходов:

$$L_{T_i} \leq T_i \leq U_{T_i}, i = \overline{1, n}, T_i \in [0, 1], \quad (5)$$

$$L_{g_i} \leq g_i \leq U_{g_i}, i = \overline{1, n}, g_i \in [0, 1], \quad (6)$$

$$L_{v_j} \leq v_j \leq U_{v_j}, j = \overline{1, m}, v_j \in [0, 1]. \quad (7)$$

Теперь, используя приведенные обозначения, формализуем возможные критерии оптимизации бюджетной политики:

- Дефицит бюджета Δ . Рассчитывается как разность между доходами и расходами бюджета:

$$\Delta = I + D - W = \sum_{i=1}^n S_i T_i + P + C \left(\sum_{i=1}^n S_i T_i g_i + E - \sum_{i=1}^n S_i T_i - P \right) - \sum_{i=1}^n S_i T_i g_i - E.$$

- Финансовая независимость региона FI . Рассчитывается как отношение собственных доходов к общим расходам:

$$FI = \frac{I'}{V} = \frac{\sum_{i=1}^n (1 - g_i) S_i T_i + P}{\sum_{i=1}^n S_i T_i g_i + E}. \quad (8)$$

- Показатель дефицитности бюджета DB . Рассчитывается как отношение дефицита бюджета к общим расходам бюджета:

$$DB = \frac{\Delta}{W} = \frac{\sum_{i=1}^n S_i T_i + P + C \left(\sum_{i=1}^n S_i T_i g_i + E - \sum_{i=1}^n S_i T_i - P \right) - \sum_{i=1}^n S_i T_i g_i - E}{\sum_{i=1}^n S_i T_i g_i + E}.$$

- Показатель дотационности бюджета TB . Рассчитывается как отношение величины полученного трансферта к собственным доходам:

$$TB = \frac{D}{I'} = \frac{C(W - I)}{\sum_{i=1}^n S_i T_i (1 - g_i) + P} = \frac{C \left(\sum_{i=1}^n S_i T_i g_i + E - \sum_{i=1}^n S_i T_i - P \right)}{\sum_{i=1}^n S_i T_i (1 - g_i) + P}.$$

- Показатель донорства региона DR . Рассчитывается как отношение налоговых отчислений к величине полученного трансферта:

$$DR = \frac{\sum_{i=1}^n S_i T_i g_i}{C \left(\sum_{i=1}^n S_i T_i g_i + E - \sum_{i=1}^n S_i T_i - P \right)}.$$

- Коэффициент бюджетной результативности региона BE . Рассчитывается как отношение собственных доходов бюджета к среднегодовой численности региона:

$$BE = \frac{\sum_{i=1}^n S_i T_i (1 - g_i) + P}{H}.$$

- Коэффициент бюджетного покрытия BC . Рассчитывается как отношение общих доходов бюджета к его общим расходам:

$$BC = \frac{I' + D}{W} = \frac{\sum_{i=1}^n S_i T_i (1 - g_i) + P + C \left(\sum_{i=1}^n S_i T_i g_i + E - \sum_{i=1}^n S_i T_i - P \right)}{\sum_{i=1}^n S_i T_i g_i + E}.$$

- Коэффициент бюджетной обеспеченности населения BW . Рассчитывается как отношение величины расходов бюджета к среднегодовой численности населения региона:

$$BW = \frac{E}{H} = \frac{\sum_{j=1}^m v_j E}{H}.$$

Задача оптимизации баланса величин налоговых ставок и ставок отчислений для региональных бюджетов

Пусть $x = (x_1, \dots, x_n)$ - вектор искомых налоговых ставок, $y = (y_1, \dots, y_n)$ - вектор соответствующих ставок отчислений в бюджет высшего уровня. В качестве критерия оптимальности будем использовать коэффициент финансовой независимости бюджета (8). Целевое значение этого критерия равно 1. Учитывая это, в предложенных выше обозначениях целевую функцию задачи можно записать в виде

$$F(x, y) = \left(\frac{\sum_{i=1}^n (1 - y_i) S_i x_i + P}{\sum_{i=1}^n S_i x_i y_i + E} - 1 \right)^2. \quad (9)$$

.Тогда с математической точки зрения задача состоит в минимизации целевой функции

$$F(x, y) \rightarrow \min$$

при условии выполнения ограничений, аналогичных (5)-(6)

$$L_{T_i} \leq x_i \leq U_{T_i}, i = \overline{1, n}, x_i \in [0, 1],$$

$$L_{g_i} \leq y_i \leq U_{g_i}, i = \overline{1, n}, y_i \in [0, 1],$$

Остальные ограничения модели (1)-(4), (7) считаются выполненными, если данные заданы корректно.

Подобным образом можно ставить задачи оптимизации с другими критериями или параметрами управления.

Теперь покажем, как свести описанную выше задачу, к задаче комбинаторной оптимизации. Пусть $x = (x_1, x_2, \dots, x_n) \in X_1 \times X_2 \times \dots \times X_n$ - n -мерный вектор налоговых ставок. Множества $X_i, i = \overline{1, n}$ конечны и имеют вид

$$X_i = \left\{ x : x = L_{T_i} + j\Delta_{T_i}, j = 0, \overline{\frac{U_{T_i} - L_{T_i}}{h_{T_i}}}, x \in [0, 1] \right\},$$

где h_{T_i} - шаг дискретизации, L_{T_i} и U_{T_i} - нижнее и верхнее граничное значение для i -ой налоговой ставки.

Пусть $y = (y_1, y_2, \dots, y_n) \in Y_1 \times Y_2 \times \dots \times Y_n$ - n -мерный вектор ставок отчислений в бюджет высшего уровня. Множества $Y_i, i = \overline{1, n}$ конечны и имеют вид

$$Y_i = \left\{ y : y = L_{g_i} + j\Delta_{g_i}, j = 0, \overline{\frac{U_{g_i} - L_{g_i}}{h_{g_i}}}, y \in [0, 1] \right\},$$

где h_{g_i} - шаг дискретизации, L_{g_i} и U_{g_i} - нижнее и верхнее граничное значение для i -ой ставки отчисления.

Такой подход говорит о том, что в предложенной модели процентные ставки могут принимать только конечное число значений в рамках определенного коридора, и в процессе решения задачи оптимизации изменяться лишь на фиксированную величину – шаг. С экономической точки зрения, подобное допущение вполне допустимо, поскольку реальная налоговая ставка составляет, скорей всего, целую величину – например, 20%, а не 19,325%. К тому же, при поиске оптимальной налоговой ставки также логичнее изменять ее с целым шагом – например, на 5% или 1%.

Сама задача состоит в поиске таких векторов $x = (x_1, x_2, \dots, x_n)$ и $y = (y_1, y_2, \dots, y_n)$, для которых достигается минимальное значение целевой функции (9) при ограничениях

$$x_i \in X_i, i = \overline{1, n},$$

$$y_i \in Y_i, i = \overline{1, n}$$

Сформулированная таким образом задача является задачей комбинаторной оптимизации повышенной размерности со сложной целевой функцией, поэтому для ее решения целесообразно использовать прикладные методы комбинаторной оптимизации [Сергиенко, 2009].

Заключение

В работе предложена оптимизационная модель регионального бюджета и межбюджетных отношений. На основе этой модели формализованы важнейшие показатели управления региональным бюджетом и критерии его оптимизации, сформулирована задача поиска величины оптимальных налоговых ставок в виде задачи непрерывной и комбинаторной оптимизации.

Вообще говоря, для описания бюджетного процесса наиболее приемлемой является форма некоторого алгоритма действий (инструкций). Действительно, бюджет есть результатом последовательного принятия серии решений различными участниками бюджетного процесса. Кроме того, это рекуррентная задача, которая решается ежегодно. Практика показывает, что сложные задачи такого вида, соответствующие органы государственного управления разбивают на серию меньших подзадач и решают уже их в определенной последовательности. Очевидно, что лицам, принимающим непосредственное участие в формировании регионального бюджета, сложно учесть массу действующих факторов и их взаимосвязи, чтобы достигать оптимально сбалансированного использования ресурсов региона, распределенных функциями бюджета для достижения определенных целей. Именно поэтому, использование методов комбинаторной оптимизации при построении оптимизационной модели планирования и прогнозирования регионального бюджета является весьма перспективным подходом.

Благодарности

Статья опубликована благодаря финансовой поддержке проекта ITHEA XXI Института информационных теорий и приложений FOI ITHEA (www.ithea.org) и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Ukraine (www.aduis.com.ua).

Библиография

[Crecine, 1986] Crecine J. A computer simulation model of municipal budgeting. In: Management Science (pre-1986), Jul 1967, 13, 11, ABI/INFORM Global. – P. 786.

[Sinnuany-Stern, 1993] Sinnuany-Stern Z. A network optimization model for budget planning in multi-objective hierarchical systems // Journal of Operational Research Society. – 1993. – 44, No. 3. – P. 297-308.

[Лукша, 1997] Лукша П.О. Оптимизация бюджетного процесса на муниципальном и региональном уровне. – М: Высшая школа экономики, 1997.

[Нариньяни, 2008] http://www.raai.org/cai-08/files/cai-08_paper_167.doc

[Сергиенко, 2009] Сергиенко И.В., Гуляницкий Л.Ф., Сиренко С.И. Классификация прикладных методов комбинаторной оптимизации // Кибернетика и системный анализ. – 2009. – №5. – С. 71–83.

[Соколова, 2002] Соколова С.В. Проблемы бюджетной недостаточности муниципальных образований // Вестник СПбГУ. – 2002. – 13. – С. 34-49.

Автор

Диана Омелянчик – студентка 6-го курса факультета кибернетики Киевского национального университета имени Тараса Шевченко (г.Киев), e-mail: omelyanchikd@gmail.com.

Сфера научных интересов: математическое моделирование; комбинаторная оптимизация; математическая экономика; региональные бюджеты.

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ В РАМКАХ КОНЦЕПЦИИ "ЭЛЕКТРОННАЯ ТАМОЖНЯ"

Борис Мороз, Сергей Коноваленко

Аннотация: В работе выделена проблематика разработки и использования эффективных методов и средств обработки информации таможенного контроля. Было предложено применить в информационно-аналитических системах таможенной службы Украины математический аппарат искусственных нейронных сетей типа многослойный персептрон, где в качестве методов минимизации целевой функции использованы квазиньютоновские алгоритмы.

Ключевые слова: таможенный контроль, нейронные сети, методы минимизации.

ACM Classification Keywords: I.5 PATTERN RECOGNITION – I.5.1 Models – Neural nets.

Введение

Осуществление таможенного дела и эффективная работа таможенной службы Украины возможны лишь при использовании современных информационных технологий, методы и средства которых позволят быстрее и качественнее решать поставленные государством задачи. Концепция создания и развития комплексной многофункциональной системы "Электронная таможня" (E-customs) нацелена на укрепление таможенной безопасности Украины [Пашко, 2008], усовершенствование её информационной инфраструктуры. Реализация данной концепции осуществляется с учетом международных стандартов и с максимальным привлечением научно-технического потенциала таможенной службы Украины, что позволяет выделить такую актуальную часть исследований, как разработка и применение методов и средств обработки информации таможенного контроля.

В арсенале учёных существует немало методов и средств, которые используются для разработки систем принятия решений, классификации, распознавания, но из этой массы хочется выделить методы искусственного интеллекта на основе коннекционистского подхода. Теория нейронных сетей за последние десятилетия приобрела довольно много практических применений в разнообразных отраслях науки и техники. Так, например, их используют для распознавания лиц, изображений [Борисов, 2008], в медицине и диагностике [Ляхов и Алешин, 2010], для анализа фондового и валютного рынка [Газетдинов, 2008], для управления механизмами и роботами. Столь широкое распространение и наличие множества программных пакетов моделирования даёт нам основание предположить о возможности применения этой математической модели для нужд таможенной службы Украины. Публикаций по применению методов обработки информации таможенного контроля относительно немного и, в большинстве своём, они носят концептуальный характер, выделяя проблематику и актуальность темы. В работе [Семенко, 2008] описана возможность применения теории нечётких множеств с целью анализа рисков, что позволило, применяя функции принадлежности $\mu(x)$, оперировать более гибко с расчетом степени риска. Нам, в свою очередь, требуется обеспечить возможность обучения и адаптации системы, что направляет наш выбор в пользу многослойных нейронных сетей.

Постановка задачи

Целью статьи является рассмотрение вопросов теоретического и практического применения математического аппарата нейросетевого моделирования для системы анализа рисков нарушения таможенного законодательства, в связи с чем возникает необходимость в решении следующих задач:

1. обработать предметную область "информация таможенного контроля";
2. сформировать архитектуру нейронной сети;
3. провести анализ методов обучения и качества работы нейросетевого классификатора;
4. на основе проведенных экспериментов выбрать лучшее решение задачи распознавания рисков.

E-customs – автоматизация таможенных процедур

Тенденции мировой торговли и развитие внешнеэкономической деятельности диктуют свои условия по упрощению таможенных процедур во время ввоза и вывоза товаров, уменьшению рисков нарушения экономической безопасности. Эти условия реализуются путём автоматизации:

1. контроля по перемещению товаров и транспортных средств согласно заявленному таможенному режиму;
2. процессов таможенного оформления товаров, транспортных средств и граждан;
3. учёта субъектов внешнеэкономической деятельности;
4. процессов выявления и учёта нарушений таможенных правил;
5. систем сбора, обработки и накопления информации для формирования таможенной статистики;
6. интеграции системы анализа рисков в автоматизированные системы таможенного оформления и т.д..

Поддержка развития внешнеэкономической торговли тесно связана с упрощением таможенных процедур, но, в тоже время, это должно согласовываться с защитой экономических интересов Украины. В связи с этим, акценты реализации таможенного дела смещаются в пользу применения методов управления рисками, использования новых технологий неразрушающего контроля, тесного сотрудничества с таможенными службами других стран. Это даёт возможность оптимально использовать имеющиеся в наличии ресурсы таможенной службы и производить контроль таможенного оформления на тех участках, где существует наибольший риск нарушений, позволяя основной массе товаров и физическим лицам сравнительно свободно проходить таможенный контроль.

Рассмотрим более детально систему анализа и управления рисками, где ключевым понятием есть индикатор риска. Индикатор риска – это определенный критерий, который используется для выявления потенциальных нарушений таможенного законодательства. В табл. 1 приведены некоторые примеры индикаторов риска, сгруппированные в такие сферы риска, как характер и цель нарушения таможенного законодательства.

Эта система, используя введенные данные, должна осуществлять оценку риска по декларации либо ситуации во время таможенного оформления. Если возник риск, то система выдаёт рекомендации по применению необходимых мероприятий, направленных на минимизацию ущерба от потенциального нарушения таможенного законодательства.

Таблица 1. Некоторые виды характера и причин нарушения таможенного законодательства

Возможный характер нарушения	Неправильная классификация товара
	Неправдивые данные о происхождении товара
	Занижение/завышение таможенной стоимости товара
	Недекларирование товара
	Контрабанда/сокрытие от таможенного контроля
	Неправильные данные о причине получения льгот по оплате таможенных платежей
	Неправильные данные о характеристике/свойствах товара
	Декларирование товара не своим именем

Возможная цель нарушения	Уклонение от уплаты/уменьшения размера пошлины
	Уклонение от уплаты/уменьшения размера особого вида пошлины (антидемпинговая, специальная, компенсационная)
	Возмещение НДС
	Уклонение от уплаты/уменьшения размера акцизного сбора
	Использование таможенных льгот
	Уклонение от нетарифных ограничений

Выделим из предметной области таможенного контроля некоторые идентификационные характеристики, используемые E-customs, по значениям которых возможно будет классифицировать риск как "Высокий", "Умеренный", "Низкий":

1. x_1 – страна происхождения товара;
2. x_2 – код товара;
3. x_3 – таможенная стоимость;
4. x_4 – количество товара;
5. x_5 – вес товара;
6. x_6 – фактурная стоимость товара;
7. x_7 – разница брутто и нетто товара;
8. x_8 – история участника внешнеэкономической деятельности.

Данные идентификационные признаки и будут использоваться в качестве информационного вектора, поданного на вход системы распознавания.

Применение нейронных сетей для решения задач E-customs

Нейронные сети обладают свойствами, которые необходимы для решения практических задач: возможность обучаться и адаптироваться к изменяющимся условиям, обобщать накопленные знания, универсально аппроксимировать функцию от нескольких переменных, параллельно обрабатывать информацию, обладать устойчивостью к ошибкам за счет множества межнейронных связей [Bishop, 1996].

Для решения задачи классификации при помощи нейронных сетей необходимо определенным образом подготовить входные данные [Swingler, 1996], определить количество слоев и, соответственно, количество нейронов в них. У нас есть в наличии входной вектор $x = [x_0, x_1, \dots, x_i]^T$, который имеет в своей структуре разнородные типы данных, нуждающихся в нормировке и кодировании, дабы наша сеть могла с ним работать. Для примера был взят товар такой категории, как "микроконтроллеры и микрокомпьютеры" (8542 21 50 00 – номер по классификатору). Пример процесса формирования входного вектора приведен в табл. 2. Вектор желаемых результатов $d = [d_1, d_2, \dots, d_j]^T$ имеет размерность три, где каждый элемент вектора принимает значение 0 либо 1. Причем, если один из элементов установлен в 1, то все остальные равны 0, что и будет соответствовать одному из уровней риска {"Низкий", "Умеренный", "Высокий"}.

Таблица 2. Формирование входного вектора признаков

№	Идентификационные характеристики	Тип данных	Принимаемые значения
X ₀	Страна происхождения товара	строковый	Оффшорные зоны
			Страны ЕС
			Страны ЕЭП
X ₁	Код товара	целый	В соответствии с классификатором УКТВЭД
X ₂	Таможенная стоимость	вещественный	В соответствии с таможенной декларацией
X ₃	Количество товара	целый	Количество штук либо партий поставки
X ₄	Вес товара	вещественный	Вес единицы товара либо партии поставки
X ₅	Фактурная стоимость товара	вещественный	В соответствии с таможенной декларацией
X ₆	Разница брутто и нетто товара	вещественный	не более 5%
			от 5% до 8%
			более 8%
X ₇	История участника внешнеэкономической деятельности	строковый	Черный список
			Серый список
			Белый список

Компоненты входного вектора разбиваются на три диапазона и кодируются. Например, признак X₆ имеет значения {не более 5%; от 5% до 8%; более 8%}, которые после преобразования принимают вид {0; 0,5; 1}. В данном случае выход сети должен быть равен {0; 0; 1}, что интерпретируется как высокий уровень риска, поскольку присутствует недопустимая разница между брутто и нетто товара.

Сформируем структуру нейронной сети. В качестве архитектуры сети выберем прямосвязный многослойный перцептрон. Обычно он состоит из множества сенсорных элементов, которые образуют входной (рецепторный) слой сети (input layer), одного или несколько скрытых слоев вычислительных нейронов (hidden layer) и одного выходного слоя нейронов (output layer). Так как количество нейронов во входном и выходном слое обусловлено размерностью соответствующих векторов, то входной слой состоит из 8 нейронов-рецепторов, а выходной слой содержит 3 нейрона. Назначим каждому нейрону выходного слоя один из трех уровней риска. Количество скрытых слоев и число нейронов в них уточняется в процессе экспериментальных исследований.

В качестве функции активации нейронов скрытых и выходного слоёв будем использовать униполярный сигмоид:

$$f(s) = \frac{1}{1 + e^{-s}}, \quad (1)$$

где s – состояние нейрона.

Алгоритм обучения многослойного персептрона носит пошаговый характер, где на каждой итерации подаются все элементы обучающей последовательности (x, d) и на выходе сети вычисляется значение целевой функции (ошибки выходного слоя с P нейронами):

$$E(w) = \frac{1}{P} \sum_{j=1}^P (d_j - y_j)^2, \quad (2)$$

где d_j – желаемое выходное значение при входном векторе X^n , y_j – сигнал j -го нейрона выходного слоя.

Если значение $E(w)$ больше установленного порога, то алгоритм изменяет веса нейронов выходного и скрытых слоев нейронной сети. Эта процедура повторяется для всех векторов обучающей выборки. Обучение происходит до тех пор, пока значение целевой функции не будет соответствовать нужной точности. Целевая функция $E(w)$, а также ее градиент используются для корректировки весов нейронной сети $w_{(k+1)}$:

$$w_{(k+1)} = w_{(k)} + \eta p(w), \quad (3)$$

где η – коэффициент обучения, $p(w)$ – вектор направления целевой функции.

Более детально с процедурой обучения нейронных сетей можно ознакомиться в работах [Наукин, 1998] и [Каллан, 2001]. Одним из важных аспектов обучения сети является выбор метода минимизации целевой функции. Традиционные градиентные методы не всегда эффективно решают задачу глобальной оптимизации, останавливаясь иногда на локальных минимумах, поэтому будем использовать методы, которые учитывают вторые производные целевой функции. К таковым относятся методы, реализующие ньютоновскую стратегию, такие как метод Левенберга-Марквардта (LM) и Квазиньютоновский метод (BFGS) [Осовский, 2002]:

$$w_{(k+1)} = w_{(k)} - H_k^{-1} g_k, \quad (4)$$

где H_k – матрица Гессе, g_k – вектор градиента,

Следует отметить, что эти методы используют аппроксимированное значение гессиана $H(w)$.

Экспериментальная нейронная сеть моделировалась в среде MATLAB 7. Для сравнительного анализа исследовались сети с одним и двумя скрытыми слоями. В качестве критерия останова обучения использовалось значение ошибки обучения равное 0,01 либо, как показатель, слишком малое ее изменение при значительном увеличении количества эпох (более 300, при $\max=700$). Для обучения нейронной сети использовалось обучающее множество количеством 950 векторов.

Для определения качества распознавания сети использовалась тестовая выборка объемом 530 векторов, отличных от обучающей выборки. Получаемый при этом выходной вектор $Y = [y_1, y_2, \dots, y_j]^T$ подавался на вход функции `compnet()`, которая преобразовывала его таким образом, что максимальное значение вектора заменялось 1, а все остальные устанавливались в 0. Во время тестирования нейронных сетей производился подсчет неправильно распознанных образов, результаты которого приведены в табл. 3-4.

В табл. 3 приведены результаты работы сети с одним скрытым слоем, количество нейронов в котором варьировалось от 5 до 20 с шагом 3.

Табл. 4 содержит результаты работы сети с двумя скрытыми слоями, где количество нейронов в первом скрытом слое варьировалось от 5 до 20 нейронов с шагом 3, а во втором – от 5 до 20 с шагом 1. Каждый из 6 экспериментов проводился с постоянным количеством нейронов первого скрытого слоя и переменным значением второго слоя. Значение ошибки распознавания в пределах одного эксперимента имело свойство как уменьшаться, так и увеличиваться, образуя переменный интервал.

Таблица 3. Результаты распознавания нейронной сети с одним скрытым слоем (in-[h1]-out)

№	Количество нейронов в скрытом слое h_1	Количество неправильно распознанных образов, %	
		Метод <i>LM</i>	Метод <i>BFGS</i>
1	5	19,5	28,6
2	8	14,2	27,2
3	11	6,7	23,4
4	14	4,1	18,7
5	17	2,2	15,9
6	<u>20</u>	<u>1,4</u>	9,3

Таблица 4. Результаты распознавания нейронной сети с двумя скрытыми слоями (in-[h1-h2]-out)

№	Количество нейронов в 1-ом скрытом слое h_1	Количество нейронов во 2-ом скрытом слое h_2 (диапазон)	Количество неправильно распознанных образов (интервал значений), %	
			Метод <i>LM</i>	Метод <i>BFGS</i>
1	5	[5 .. 20]	[14,1 .. 2,5]	[27,6 .. 20,2]
2	8	[5 .. 20]	[8,3 .. 6,9]	[27,2 .. 9,7 .. 19,1]
3	11	[5 .. 20]	[3,5 .. 2,2 .. 6,3]	[26,8 .. 11,4 .. 17,3]
4	14	[5 .. 20]	[4,5 .. 1,5 .. 4,2]	[5,1 .. 7,4 .. 4,1]
5	<u>17</u>	[5 .. <u>11</u> .. 20]	[2,2 .. <u>1,2</u> .. 2,9]	[7,6 .. 4,4 .. 6,2]
6	20	[5 .. 20]	[1,4 .. 4,7 .. 3,1]	[4,6 .. 3,7 .. 5,2]

Проведя по 6 экспериментов для различных архитектур многослойного персептрона, выявлено, что Квазиньютоновский метод (BFGS) минимизации целевой функции уступает методу Левенберга-Марквардта (LM) по времени обучения нейронной сети и качеству распознавания тестовой выборки.

На основе результатов экспериментов можно выделить оптимальные модели нейронной сети для одного и двух скрытых слоев:

1. архитектура с одним скрытым слоем – **8-[20]-3**, метод обучения LM, ошибка распознавания 1,4 %;
2. архитектура с двумя скрытыми слоями – **8-[17-11]-3**, метод обучения LM, ошибка распознавания 1,2 %.

В общем случае для данного набора исходных данных предпочтительней использовать первую модель, так как она содержит меньшее количество межнейронных связей, тем самым облегчая процесс обучения нейронной сети.

Заключение

В результате проделанной работы были разработаны модели идентификации рисков нарушения таможенного законодательства на базе нейронной сети типа многослойный персептрон. На основе проделанных опытов и сравнительного анализа качества распознавания входных образов, выделена оптимальная архитектура нейронной сети.

Разработанная модель показала, что применение нейронных сетей в качестве методов классификации информации таможенного контроля, может быть полезным инструментом в развитии многофункциональной комплексной системы "Электронная таможня".

Дальнейшие исследования следует посвятить изучению эффективных методов расчета оптимальной архитектуры многослойного перцептрона, а также изучению механизма улучшения возможностей нейроклассификатора за счет совместного использования с нечеткой логикой, применением генетических алгоритмов.

Благодарности

Работа опубликована при финансовой помощи из проекта ITHEA XXI Института Информационных теорий и Приложений FOI ITHEA (www.ithea.org) и Ассоциации Создателей и Пользователей Интеллектуальных Систем АСПИС Украина (www.aduis.com.ua)

Библиография

- [Пашко, 2008] Пашко П.В. Основы митної справи в Україні: Підручник – К.: Знання, 2008. – 652 с.
- [Борисов, 2008] Евгений Борисов. Использование искусственных нейронных сетей для классификации черно-белых изображений // Кибернетика и системный анализ. – 2008. – №2. – с. 184-187.
- [Ляхов и Алешин, 2010] Ляхов А.Л., Алешин С.П. Нейросетевая модель формулы сбалансированного питания // Искусственный интеллект. – 2010. – №2. – с. 131-139.
- [Газетдинов, 2008] Газетдинов В.А. Нейросетевая система анализа и прогнозирования процессов на рынке недвижимости // Искусственный интеллект. – 2008. – №2. – с. 50-57.
- [Семенко, 2008] Семенко, О. М.-М. Підвищення ефективності роботи по протидії контрабанді та порушенням митних правил, митного контролю за рахунок впровадження автоматизованої системи аналізу та управління митними ризиками на основі Fuzzy - технології: до вивчення дисципліни / О. М.-М. Семенко. – К.: LAT & K, 2008. – 238 с.
- [Bishop, 1996] Christopher M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press, USA; 1 edition, 1996, p. 504.
- [Swingler, 1996] [Kevin Swingler](#). Applying Neural Networks: A Practical Guide. Morgan Kaufmann; Pap/Disk edition, 1996, p. 303.
- [Haykin, 1998] Simon Haykin. Neural Networks: A Comprehensive Foundation. Prentice Hall; 2 edition, 1998, p. 842.
- [Каллан, 2001] Роберт Каллан. Основные концепции нейронных сетей: Пер. с англ. / Роберт Каллан. – М.: Издательский дом "Вильямс", 2001. – 287 с.
- [Осовский, 2002] С. Осовский. Нейронные сети для обработки информации: Пер. с польского И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344 с.

Сведения об авторах



Борис Мороз – доктор технических наук, профессор, советник таможенной службы I ранга, декан факультета информационных и транспортных систем и технологий Академии таможенной службы Украины, улица Дзержинского, 2/4, Днепропетровск, 49000, Украина; e-mail: realtime07@mail.ru

Основные области научных исследований: теория информации, изучение аспектов ценности и старения информации, теория нечетких множеств



Сергей Коноваленко – аспирант, инспектор таможенной службы III ранга, начальник лаборатории информационных систем и процессов в таможенном деле Академии таможенной службы Украины, улица Дзержинского, 2/4, Днепропетровск, 49000, Украина; e-mail: customslab@rambler.ru

Основные области научных исследований: методы искусственного интеллекта, нейронные сети, вычислительные алгоритмы природной мотивации

НЕЙРОСЕТЕВЫЕ АЛГОРИТМЫ ПОИСКА РЕШЕНИЯ ОБРАТНОЙ ЗАДАЧИ КИНЕМАТИКИ РОБОТА ВЕРТИКАЛЬНОГО ПЕРЕМЕЩЕНИЯ (РВП)

Тимофеев Адиль Васильевич, Титов Виктор Викторович

Abstract: Рассматривается вариант использования нейронных сетей в системе управления среднего уровня робота вертикального перемещения. Предлагается и обсуждается структура управления движением двуногого присосочного робота с использованием нейросетевого приближения решения обратной задачи кинематики при планировании движений.

Keywords: Робот вертикального перемещения, обратная задача кинематики, планирование движений, нейросетевая аппроксимация, стохастический поиск.

ACM Classification Keywords: I.2.9 Robotics: Kinematics and dynamics.

Введение

Мобильные роботы в настоящее время используются как в промышленности, так и в местах и средах, труднодоступных или опасных для человека (например, в космическом пространстве, под водой, а также в условиях высокой температуры и радиации). Одним из относительно новых направлений в развитии мобильной робототехники является создание роботов вертикального перемещения (РВП). Подобные роботы способны выполнять операции различного рода, находясь на поверхности с произвольным углом наклона.

Данное направление робототехники начало активно развиваться в 90-х годах в таких странах, как Япония, США, Англия, Германия, а также в России. Вначале 90-х были выпущены многочисленные статьи, посвященные созданию или описанию уже действующих прототипов роботов вертикального передвижения. Среди российских разработок можно выделить транспортный робот вертикального передвижения (ТРВП), описанный в [Градецкий, 1997].

Исследования в этом направлении используют для достижения поставленных задач разнообразные механизмы, материалы и принципы движения. В частности, широко используются вакуумные механизмы (присоски), новейшие материалы в аппаратах крепления к поверхности (как правило, это различного рода адгезионные поверхности с микроструктурным и наноструктурным материалом) и принципы передвижения, заимствованные из природы (*bio-inspired hardware*).

Разработанный в ЦНИИ РТК робот вертикального перемещения (рис.1) имеет пять степеней подвижности и использует две вакуумные присоски (стопы) для перемещения по плоским вертикальным и поверхностям произвольной ориентации.

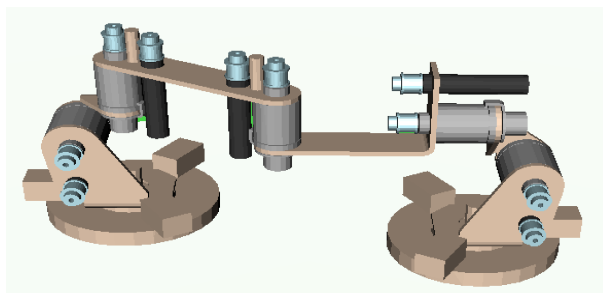


Рис. 1. Стопоходящий робот вертикального перемещения

Постановка задачи

Верхний уровень управления мобильными роботами, отвечающий за тактическое планирование движений и перемещений, как правило, берет на себя нагрузку по формированию формы и параметров траектории движений сочленений робота с целью обеспечения его общего перемещения в пространстве. При планировании тактический уровень управления должен учитывать как специфику конфигурации робота, так и особенности непосредственной окружающей среды робота (локальная навигация и обход препятствий).

Для выполнения такого планирования алгоритм должен уметь рассчитывать положения сочленений в зависимости от обобщенных координат механизма, а также уметь решать обратную данной задачу за минимально возможное время.

Другим аспектом работы данного уровня управления является генерация траектории движения механизма в среде с препятствиями за ограниченное время.

Таким образом, при разработке принципов передвижения рассматриваемого робота решались две перечисленные задачи тактического уровня управления:

- быстрое решение обратной задачи кинематики;
- метод генерации траектории движения сочленений при наличии препятствий и кинематических ограничений

Решение обратной задачи кинематики

В ходе экспериментальных работ и исследований по моделированию была подтверждена необходимость в решении обратной задачи кинематики данного механизма, применяемая в задачах контроля положения, планирования и калибровки. Из множества подходов к решению данной задачи ([Тимофеев,1999], [Тимофеев,1988], [Фу, 1989]) был выбран подход, связанный с численной минимизацией функционала

$$Z(q) = Q_{зад} - F(q) \quad (1)$$

где q – углы шарниров, $F(q)$ – решение прямой задачи кинематики, $Q_{зад}$ – целевое положение и ориентация в декартовых координатах.

Как известно, при использовании градиентных и других методов поиска минимумов этого функционала имеются известные проблемы, связанные со скоростью сходимости алгоритмов оптимизации функционала (1), зависящие от начальных приближений q_0 .

В качестве альтернативы был рассмотрен подход, основанный на нейросетевой аппроксимацией обратной задачи кинематики. Данный подход столкнулся с рядом трудностей, связанных с невозможностью достаточно точно аппроксимировать все пространство допустимых конфигураций робота нейронной сетью приемлемой сложности. Время обучения сети и генерации обучающих выборок в рамках такого подхода оказалось велико.

Однако использование нейросетевого приближения для последующего численного решения задачи минимизации функционала (1) оказалось более выгодным по сравнению с табличным хранением начальных приближений. Помимо этого был исследован подход, при котором нейросеть приемлемой сложности (размерности) обучается на заранее заданной (или спланированной) траектории движения звеньев механизма с учетом конструктивных ограничений на углы поворота шарниров при этом движении.

Совокупность таких предобученных нейронных сетей (рис.2) является решением обратной задачи кинематики для всего пространства конфигураций. Она может быть использована как с алгоритмом численного уточнения, так и без него.



Рис. 2. Схема функционирования множества сетей

Решения обратной и прямой задачи кинематики играют важную роль при планировании движений любого робота. Как показали модельные эксперименты, решение обратной задачи кинематики при плоскостном хождении робота смысла не имеют ввиду того, что те классы препятствий (ступени не выше 5 см), которые робот способен преодолевать, легко преодолеваются за счет трех механических датчиков определения углов наклона присоски к поверхности (они же являются детекторами касания поверхности). При переходе на стену обратная задача кинематики, решаемая с помощью нейросетевого приближения с численным уточнением, в дальнейшем применялась при генерации промежуточных точек в алгоритме поиска траектории перехода на стену, а также для визуализации положения звеньев механизма при ручном управлении в сложно проходимых местах.

Стохастический алгоритм планирования траектории

Для случаев движения робота в среде с препятствиями или перехода на другие поверхности был разработан алгоритм стохастического планирования траектории движения сочленений (являющийся модификацией алгоритмов, предложенных в [Siciliano, 2009]). Одной из особенностей алгоритма является использование пошагового приближения к желаемой конфигурации робота из начальной конфигурации при стохастическом выборе величины и направления шага. Для величины и направления шага было использовано нормальное распределение. При этом математическое ожидание величины шага изменялось пропорционально обобщенному расстоянию до целевой конфигурации, что обеспечивало подход к целевой конфигурации лишь с незначительными колебаниями. Направление шага на каждой итерации также выбиралось, исходя из нормального распределения по каждой из компонент вектора обобщенных координат механизма. При этом среднеквадратичное отклонение целенаправленно уменьшалось в случае движения в направлении, близком к направлению на целевую конфигурацию. Такой подход позволяет отнести разработанный алгоритм к алгоритмам целенаправленного (*goal directed*) стохастического поиска. В случае столкновения с препятствием в ходе планирования движения среднеквадратичное отклонение направления и математическое ожидание величины шага получают дополнительное скачкообразное увеличение, которое затем снижается в ходе планирования. Такое увеличение позволяет алгоритму быстрее находить обход/выход из препятствия. Стартуя в начальной точке, алгоритм выбирает несколько точек вокруг текущей по нормальному распределению с небольшой дисперсией. Затем каждая сгенерированная точка проверялась на столкновение со стеной или выход за пределы углов механизма, а также на возможность перехода в нее из предыдущей конфигурации. Если точка проверку не проходила, то выбрасывалась из списка. Из оставшихся точек в качестве заносимой в траекторию выбирается та, которая наиболее близка к целевой конфигурации.

Кроме того, была введена дополнительная небольшая вероятность выбора из сформированного списка произвольной точки, а не наиболее близкой к целевой конфигурации. Также был рассмотрен вариант с

присвоением каждой точке списка вероятности быть выбранной, пропорциональной ее близости к целевой конфигурации (данный подход показал более быструю сходимость в сравнении с первым вариантом).

Таким образом, обеспечивалось плавное продвижение в направлении целевой конфигурации. Время планирования перехода на стену для одной ноги в среднем составило 25 секунд, что в среднем почти на 60% быстрее, чем аналогично построенное планирование, но без адаптации параметров случайного распределения (в среднем 60 секунд).

Заключение

Реализация алгоритма поиска решения обратной задачи кинематики с применением нейронных сетей, позволило в значительной степени увеличить быстродействие других алгоритмов, обрабатывающих данные о положении звеньев и шарниров, и увеличило общее качество функционирования системы.

Использование целенаправленного стохастического планирования позволило увеличить скорость работы алгоритма на открытых (без препятствий) зонах пространства состояний. Однако целенаправленность ухудшает поведение алгоритма в зонах пространства состояний с вогнутыми препятствиями. Тем не менее, заданная конфигурация робота позволяет использовать данный алгоритм достаточно эффективно.

Решение данных задач практически полностью покрывает потребности тактического уровня управления рассматриваемого робота, что позволяет ему свободно двигаться и маневрировать в любых направлениях, используя циклически повторяющиеся движения.

В дальнейшем планируется рассмотреть возможность автоматического синтеза походки всего робота, как, например, это было сделано в работе [Dittrich, 1998].

Работа выполнена при частичной поддержке гранта РГНФ–ГФЕН Китая № 10–08–91159–ГФЕН–а “Исследование научных проблем интеллектуального управления магистральными транспортными средствами и мобильными роботами”.

Библиографический список

- [Градецкий, 1997] Градецкий В.Г., Рачков М.Ю., Роботы вертикального перемещения, М.: Тип. Мин. Образования РФ, 1997, 223с.
- [Тимофеев, 1999] Тимофеев А.В., Экало Ю.В. Системы цифрового и адаптивного управления роботом: учебн. пособие. СПб: Изд-во СПбГУ, 1999. 248 с.: ил. ISBN 5-288-01081-1.
- [Тимофеев, 1988] Тимофеев А. В. Адаптивные робототехнические комплексы. // Л.: Машиностроение. 1988. 332 с.
- [Фу, 1989] Фу К., Гонсалес Р., Ли К. Робототехника: пер. с англ. – М.: Мир, 1989. – 624 с., ил. ISBN 5-03-000805-5
- [Siciliano, 2009] B. Siciliano, L. Sciacivico, L. Villani, G. Oriolo Robotics Modelling, Planning and Control 2009 Springer-Verlag London Limited ISBN 978-1-84628-641-4
- [Dittrich, 1998] Peter Dittrich, Andreas Bürgel, Wolfgang Banzhaf, Learning to move a robot with random morphology, CONFERENCE PAPER, Conf. First European Workshop on Evolutionary Robotics 1998, p. 165--178

Информация об авторах

Тимофеев Адиль Васильевич – заведующий лабораторией информационных технологий в управлении и робототехнике Санкт-Петербургского института информатики и автоматизации Российской академии наук, доктор технических наук, профессор, Заслуженный деятель науки РФ, 199178, Россия, Санкт-Петербург, 14-я линия, д. 39, СПИИРАН, tav@iias.spb.ru

Титов Виктор Викторович – аспирант Санкт-Петербургского института информатики и автоматизации Российской академии наук, 199178, Россия, Санкт-Петербург, 14-я линия, д. 39, СПИИРАН, victortitov2005@mail.ru

РАСПРЕДЕЛЕННАЯ ИНФОРМАЦИОННАЯ СИСТЕМА АГРОМОНИТОРИНГА

Андрей Шелестов, Наталья Морзе, Ольга Куссуль, Юлия Грипич

Резюме: В работе рассмотрена распределенная система агромониторинга, созданная в Украине для пользователей уровня министерства и отдельного фермерского хозяйства. Мониторинг основывается на спутниковых данных и продуктах среднего и высокого разрешения. Система включает геопортал с Web-интерфейсом, а также упрощенную настольную версию, основанную на использовании ГИС-системы с расширенными функциональными возможностями автоматической загрузки спутниковых данных и бизнес-аналитики. Система построена на базе программного обеспечения с открытым кодом, удовлетворяющего стандартам OGC для обмена геопространственной информацией.

Ключевые слова: агромониторинг, геопортал, ГИС, OGC, спутниковые данные, распределенная система, Web-интерфейс, MODIS, геопространственный интеллект, Openlayers, AJAX.

ACM Classification Keywords: D.2.12 [Software Engineering] Interoperability; Information Systems; H.1.1 [Models and Principles] Systems and Information Theory; H.3.5 [Information Storage and Retrieval] Online Information Services; I.4.8 [Image Processing and Computer Vision] Scene Analysis - Sensor Fusion.

Введение

Современное состояние дел в мониторинге сельского хозяйства и международной кооперации в этой области состоит во внедрении инициативы сельскохозяйственного мониторинга группы GEO (Group on Earth Observations — международная программа по кооперации между правительствами стран, международными организациями и космическими агентствами, <http://earthobservations.org/index.html>). Цель GEO как межправительственной структуры состоит в повышении доступности и внедрении космических наблюдений через международную координационную деятельность, используя возрастающие возможности дистанционного зондирования Земли (ДЗЗ) для поддержки принятия решений в сложном и экологически напряженном мире. В настоящее время комитет GEO включает более 80 правительств стран (в том числе Украины), Европейскую комиссию и 56 межправительственных, международных и региональных организаций.

Сельское хозяйство является одной из девяти областей социального назначения (SBA) группы GEO для международной координации и сотрудничества. В рамках этой области сформировано научное сообщество мониторинга сельского хозяйства для разработки и реализации Глобальной системы мониторинга сельского хозяйства, которое базируется на существующих системах и международных средствах спутникового мониторинга (задача GEO AG-0703) (http://earthobservations.org/documents/cop/ag_gams/20070716).

В Европейском Союзе разрабатывается и находится на начальном этапе внедрения масштабная программа Глобального мониторинга в интересах мониторинга окружающей среды и безопасности (GMES). Программа имеет целью создание информационных систем поддержки принятия решений для учреждений Европейского Союза. Эти системы базируются на информационных сервисах мониторинга окружающей среды по нескольким направлениям, в частности мониторинга сельского хозяйства (www.gmes.info). На данный момент в рамках GMES введен сервис глобального мониторинга посевов сельскохозяйственных культур (Global Crop Monitoring), обновление карт классификации земных покровов (EUROLAND), мониторинга рационального использования культивируемых земель (Agri Environmental

Monitoring) и др. [Rembold et al, 2006]. В Объединенном исследовательском центре (JRC) Европейской Комиссии, которая предоставляет научную и техническую поддержку решений Европейской Комиссии в области сельского хозяйства и продовольственной безопасности, накоплен двадцатилетний успешный опыт использования данных ДЗЗ для решения задач сельскохозяйственной статистики и прогнозирования урожайности. С 1992 года в JRC функционирует и постоянно совершенствуется система прогнозирования урожайности AGRI4CAST.

В странах СНГ внедряются системы поддержки принятия решений в агропромышленном комплексе на основе спутникового мониторинга. В частности, в Российской Федерации Институтом космических исследований РАН создана система дистанционного мониторинга для Министерства сельского хозяйства Российской Федерации [Лупян и др., 2009]. Институтом космических исследований МОН РК по заказу Министерства сельского хозяйства Республики Казахстан создана система мониторинга сельскохозяйственных посевов зернового производства. Основной задачей этих систем является мониторинг состояния посевов и определения структуры площадей пахотных земель. Обе системы функционируют на бесплатных данных приборов MODIS [Толпин и др., 2010].

Важной современной тенденцией является постепенное удешевление данных ДЗЗ совместно с увеличением их количества и качества. В этом смысле показательной есть деятельность международного сообщества во главе с GEO, благодаря которой ценные массивы данных ДЗЗ стали бесплатными. В частности, с 2009 г. все данные программы Landsat из архива службы геологической службы США (USGS) распространяются бесплатно. В частности, данные действующих спутников среднего пространственного разрешения Landsat-5, Landsat-7 и EO-1. Данные спутников миссии LDCM также планируются распространять бесплатно. В Европейском космическом агентстве (ESA) планируется распространения всех данных спутников серии Sentinel на бесплатной основе. Таким образом, в недалеком будущем станет возможным построение систем дистанционного мониторинга на основе новых дешевых данных.

В последние годы в Украине активно развиваются методы и информационные технологии, направленные на использование методов дистанционного зондирования в решении задач различных ведомств, в том числе агропромышленного комплекса [Kussul et al, 2009; Kussul et al, 2010a; Kussul et al, 2010b; Popov et al, 2008]. В данной статье описывается подход к созданию распределенной информационной системы агромониторинга, которая находит свое применение как на уровне министерств, так и отдельных хозяйств.

Задачи и функции системы

В рамках нескольких конкурсных проектов была разработана двухуровневая распределенная система агромониторинга на уровне министерств и отдельных хозяйств [Kussul et al, 2010; Шелестов и др, 2011]. Основное назначение системы — предоставлять информационные продукты оценки площадей сельскохозяйственных культур и мониторинга состояния посевов на регулярной основе, с привлечением как данных дистанционного зондирования земной поверхности, так и данных наземных измерений, которые обеспечивают верификацию дистанционных методов.

Система включает две составные части: подсистему уровня министерства, реализованную в виде геопортала, и подсистему уровня отдельного хозяйства, основанную на использовании геоинформационной системы с открытым кодом с возможностью получения геопространственных данных и спутниковой информации из системы уровня министерства.

Система реализована с использованием программного обеспечения с открытым кодом. Спутниковая и другая геопространственная информация предоставляется по стандартам OGC и может использоваться в других мониторинговых и ГИС-системах.

Доступ к системе через Web-интерфейс обеспечивается через Web-сайт проекта по адресу <http://agro.ikd.kiev.ua> (рис. 1).

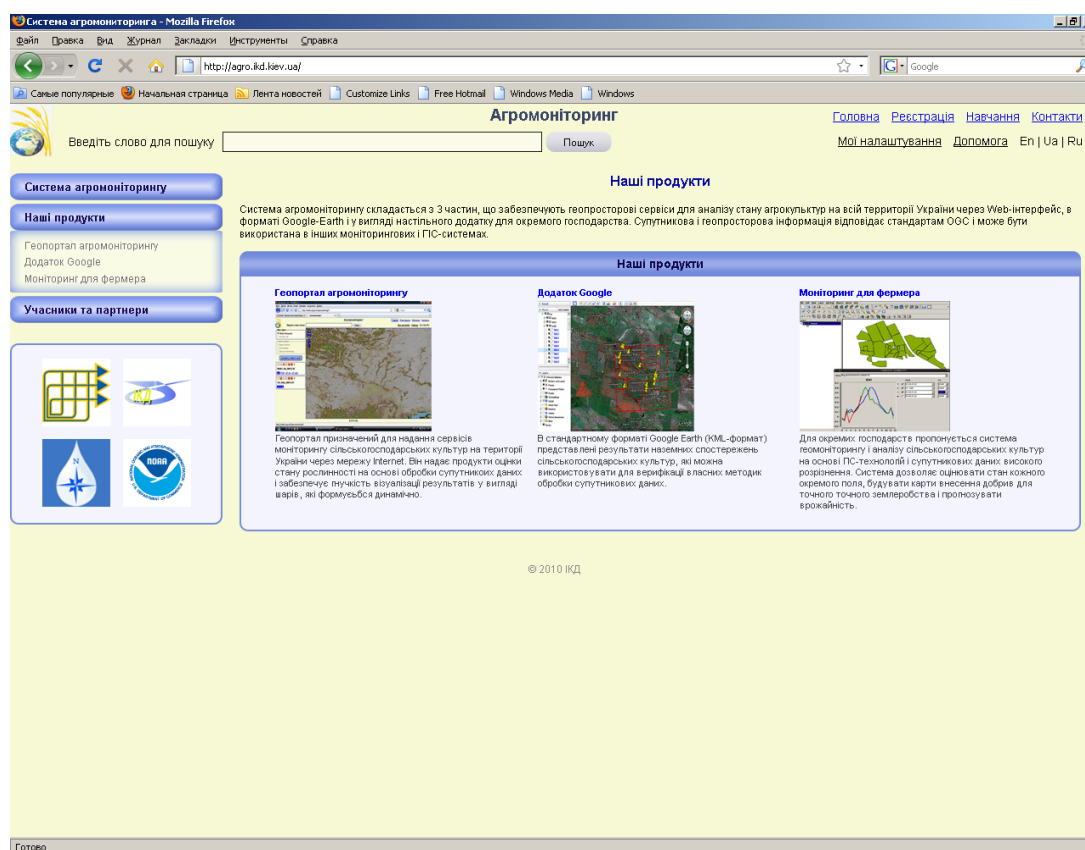


Рис. 1. Главная страница Web-сайта проекта

Геопортал агромоніторингу

Основная концепция. Предлагаемая распределенная система предоставляет информационные продукты оценки площадей сельскохозяйственных культур и мониторинга состояния посевов на регулярной основе, с привлечением как данных дистанционного зондирования земной поверхности, так и данных наземных измерений, которые должны обеспечить верификацию дистанционных методов через Web-интерфейс.

Доступ к геопорталу, основному компоненту информационной системы, и предоставляемым ею продуктам обеспечивается по адресу <http://agro.ikd.kiev.ua> (рис. 2).

К основным преимуществам геопортала следует отнести следующие.

- Отсутствие требований к использованию специализированного программного обеспечения, поскольку геопортал предназначен для использования через браузер;
- Автоматическая загрузка новых продуктов и размещения их в базе данных. Среди программных компонентов геопортала есть целый ряд специализированных модулей, предназначенных для автоматического сканирования удаленных архивов данных наблюдения Земли, их получение и размещение в базе данных. Возможность гибкой настройки динамических слоев и продуктов на базе существующей информации.
- Используемые технологии на основе существующей в базе данных информации позволяют динамически формировать различные слои геопространственного изображения.
- Возможность работы в заданном временном диапазоне.
- Средства пользовательского интерфейса позволяют в интерактивном режиме генерировать запросы на получение информации из БД для заданного периода времени.

– Использование только лицензионно чистого программного обеспечения. Все проектные решения базируются на использовании бесплатного программного обеспечения с открытым кодом.

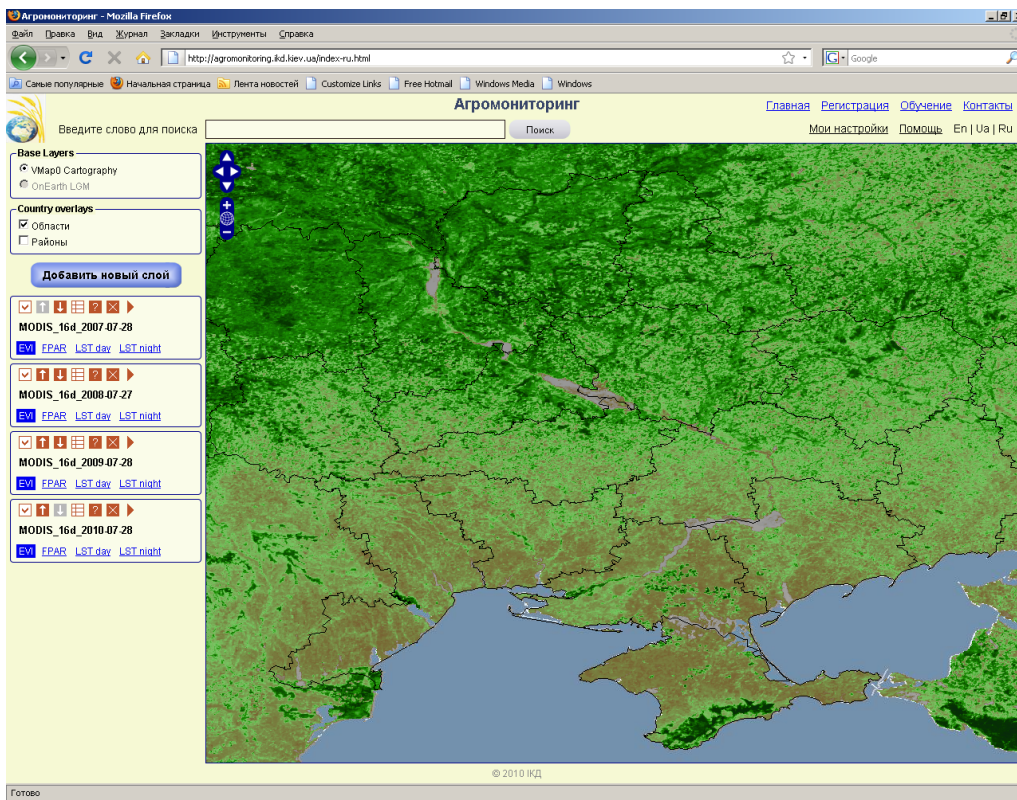


Рис. 2. Геопортал агромониторинга

Следует отметить, что в основу разработанного порталного решения положена архитектура «клиент-сервер» на основе шаблона «тонкого» клиента. Серверные компоненты реализованы на языке программирования Python, предоставляющие возможность взаимодействия с планировщиком задач и высокопроизводительной кластерной техникой. Такие возможности являются важными, поскольку позволяют эффективно решать целый ряд прикладных задач со сложным потоком выполнения задач, планировать автоматическое получение данных и т.д. При необходимости использования сложных моделей можно формировать запросы на высокопроизводительные вычисления. Архитектура системы с Web-интерфейсом приведена на рис. 3.

Средства разработки клиентской части, системные требования и функциональность. При разработке клиентской части Web-системы использовались Javascript-библиотеки Openlayers (с расширением touch.js) и JQuery. Система работает в браузерах Internet Explorer 6 и выше, Firefox, Google Chrome, Safari и Opera.

Клиентская часть представляет собой Web-страницу, в верхней части которой размещенная навигационная панель, слева — панель управления картой, а справа от нее — карта.

В списке слоев на панели управления каждый слой представлен в виде блока, где отображается название слоя (например, MODIS_16d_2009_07_01), список вариантов слоя (например, EVI, FPAR, LST day, LST night), один из которых (по умолчанию выбран первый), а также панель инструментов (в виде набора пиктограмм).

Панель инструментов имеет такие функции:

- показать/скрыть слой;
- переместить слой вверх (меняется порядок в списке слоев и на карте);
- переместить слой вниз;
- показать/скрыть легенду;
- справка;
- удалить слой;
- выбор даты (по щелчку на значке выбора даты появляется панель со списком дат. При выборе новой даты текущий слой меняет на аналогичный, соответствующий выбранной дате).

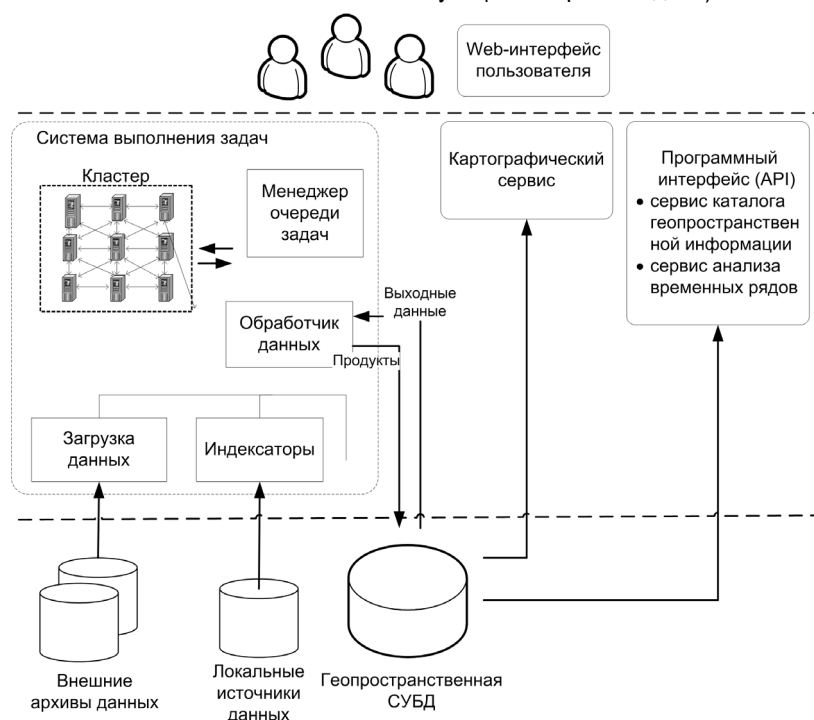


Рис. 3. Архитектура системы

После загрузки страницы на карте отображается только основной слой. В панели управления размещена кнопка добавления нового слоя. После щелчка на этой кнопке появляется панель со списком наборов данных, после выбора требуемого из которых в рабочей области отображается новый слой (например, MODIS) за указанную дату (рис. 4). Слой загружается на карту и одновременно слева в панели управления появляется соответствующий ему блок. Так можно добавлять несколько слоев, новые слои появляются поверх других, ранее сгенерированных слоев, а информация о них добавляется в начале списка слоев на панели управления.

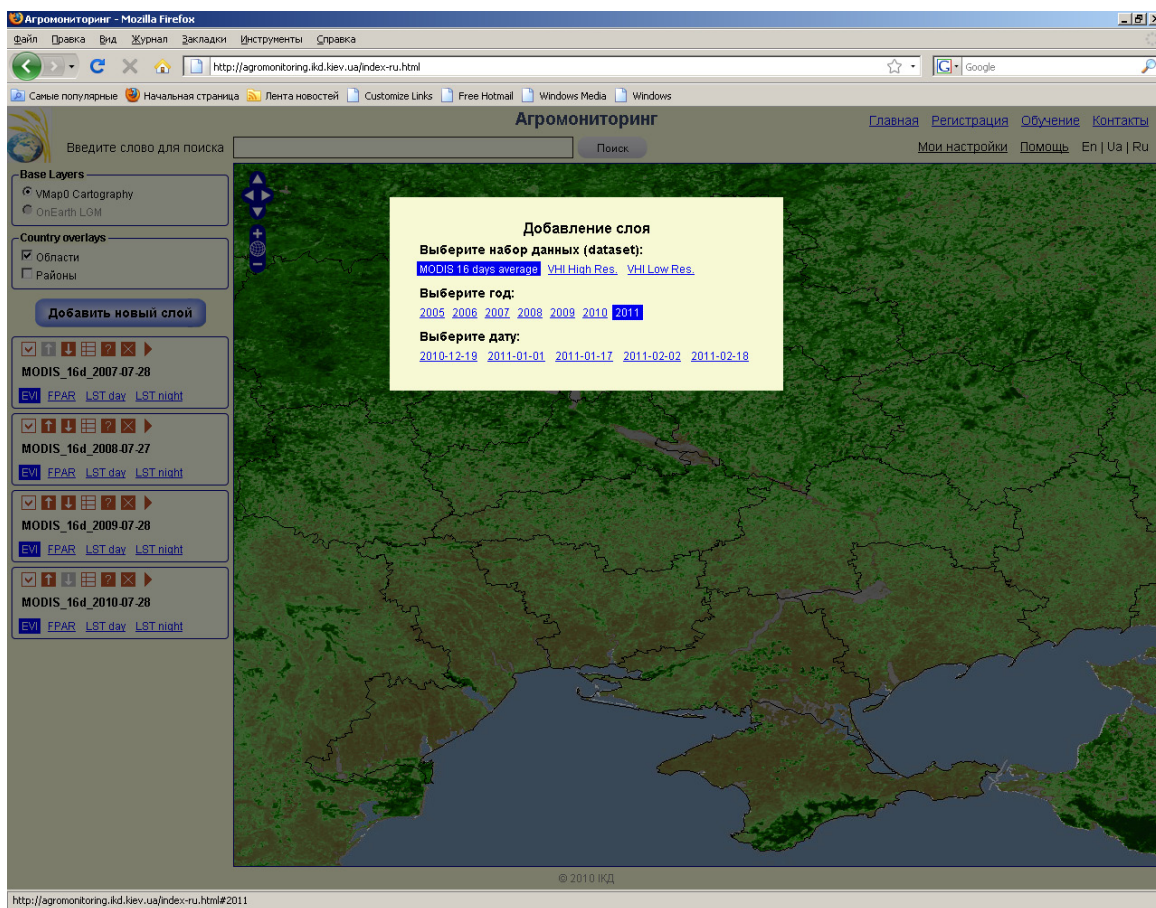


Рис. 4. Панель со списком наборов данных

Реализация клиентской части. Интерфейс реализован с помощью технологии AJAX. При работе с ним динамически загружаются отдельные части страницы, сама страница не перезагружается. Кроме указанных библиотек Javascript используется сценарий `map.js`, который и реализует основную функциональность.

При загрузке страницы сначала происходит инициализация карты (параметр `Bounds` задает ограничение на область карты для повышения скорости загрузки), добавляется базовый слой (`ol_wms_base`), карта центрируется и добавляется инициализация `touch`-интерфейса (`Touchhandler`):

Для расположения элементов на экране используется технология CSS, а также Javascript-код, который вызывается при загрузке и изменении размеров окна

Для отображения модальных блоков (например, блока добавления слоя) используются функции `addmodal` (активирует модальный блок), `addoverlay` (добавляет полупрозрачный слой, который покрывает все кроме модального блока) и `hidemodal` (скрывает все активные модальные блоки).

Запрос к программному интерфейсу сервера осуществляются с использованием подхода JSONP (который в отличие от обычного JSON позволяет реализовать кросс-доменные запросы).

Реализация клиентской части имеет достаточное быстродействие, в том числе при работе с несколькими слоями одновременно. Клиентская часть может работать как одновременно на нескольких хостах, так и на локальном компьютере или планшете с доступом к Интернет.

ПОДСИСТЕМА МОНИТОРИНГА УРОВНЯ ХОЗЯЙСТВА

Отдельным хозяйствам предоставляется упрощенная «клиентская» версия системы, построенная на основе ГИС- и мобильных технологий с использованием спутниковых данных высокого разрешения (рис. 5).

Система предоставляет удобные средства анализа состояния посевов и прогнозирования урожайности. Для отдельного хозяйства строятся :

- Точные цифровые карты отдельных полей;
- Карты внесения удобрений для точного и управляемого земледелия;
- Спутниковые данные – оперативная и объективная информация о состоянии посевов на каждом поле;
- Прогнозы урожайности.

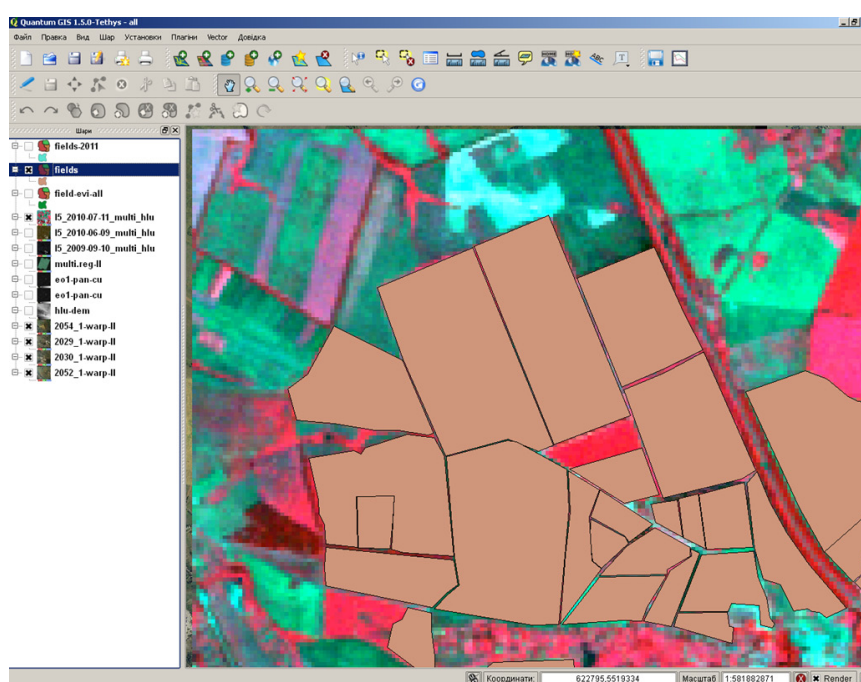


Рис. 5. Интерфейс пользователя «клиентской» версии системы агромониторинга

Архитектура подсистемы уровня отдельного хозяйства. ГИС для фермера включает в себя следующие составные части:

- Источники данных: данные из архивов пользователя (полученные из соответствующих документов пользователя, занесенные в базу данных (БД) вручную); спутниковые снимки;
- Встроенная система управления базами данных SQLite (<http://www.sqlite.org/>);
- Геоинформационная система QGIS (<http://www.qgis.org/>): дополнительный компонент к QGIS для работы с информацией из базы данных;
- Средства обработки данных: автоматизированные обработчики спутниковых изображений; административный интерфейс для работы с базой данных.

Система QGIS представляет собой программное обеспечение для работы с геопространственными данными, которое поддерживает большинство распространенных форматов данных и предоставляет программный интерфейс для реализации дополнительных компонентов, что позволяет при необходимости расширить его функциональность. QGIS является программным обеспечением с открытым кодом и распространяется в соответствии с положениями лицензии GPL. SQLite – реализация

реляционной системы управления базами данных. Благодаря легкости ее использования в качестве встроенного компонента, ее удобно использовать в качестве back-end десктопного программного обеспечения.

Процесс работы системы предусматривает агрегацию информации из разнородных источников (документация пользователя, спутниковые снимки), ее обработку, хранение в унифицированном структурированном формате, и дальнейшее предоставление информации или отчетов, сгенерированных на ее основе, через пользовательский интерфейс.

К первичным источникам информации относятся в первую очередь данные, предоставленные пользователем (исторические и текущие произведенные на полях измерения), а также спутниковые снимки.

На основе спутниковых снимков с помощью геоинформационной системы QGIS создаются векторные карты полей, которые вносятся в общую базу данных. Спутниковые снимки проходят предварительную обработку, после чего их удобно использовать для оценки параметров местности как отдельные слои QGIS. Кроме того, на основе снимков подсчитываются некоторые количественные характеристики (индексы), соответствующие отдельным полям в определенный период времени. Наличие разнообразной геопространственной информации позволяет реализовать возможности геопространственного интеллекта формирования на базе геопространственных и спутниковых данных принципиально новой информации для поддержки принятия решения.

Данные, предоставленные пользователем, упорядочиваются, и вносятся в общую базу данных с использованием интерфейса администратора для работы с БД. Интерфейс администратора предусматривает постоянный контроль целостности данных и регулярное создание резервных копий. Это защищает данные от искажения вследствие человеческих и технических ошибок.

Разработанная БД служит для унифицированного обмена информацией между компонентами системы и детально рассматривается в [Шелестов и др., 2011].

В результате все данные оказываются в общей БД (векторные, числовые) и в файлах изображений (визуальные). Данные хранятся в таком формате, чтобы система QGIS была способна отображать часть из них – векторные и растровые карты – как отдельные слои. Другая часть информации может быть прочитана и отображена с помощью специально разработанного модуля к QGIS.

Дополнительные модули. Первичная функциональность QGIS позволяет просматривать атрибуты, привязанные к объектам на карте в виде таблицы. В случае наличия большого количества числовых характеристик (например, измерений за определенный период времени) такой способ представления информации не является наглядным. Для предоставления информации в форме, удобной для пользователя, был разработан модуль, расширяющий функциональность QGIS. Данное программное обеспечение (ПО) предоставляет специальный программный интерфейс для разработки и встраивания в систему дополнительных модулей, который поддерживает языки программирования C++ и Python. В данном случае использовался язык Python, а также графическая библиотека QT.

Одно из расширений функциональности системы QGIS представляет собой инструмент для построения и сравнения графиков индексов по отдельным полям за выбранный период времени. С его помощью на основе информации из БД можно выбрать любое из имеющихся полей, период времени, и сгенерировать требуемые графики (рис. 6).

Для отображения графиков используется расширения для графической библиотеки QT - библиотека QWT. Данные для графиков получают непосредственно из общей базы данных с помощью запроса на языке SQL.

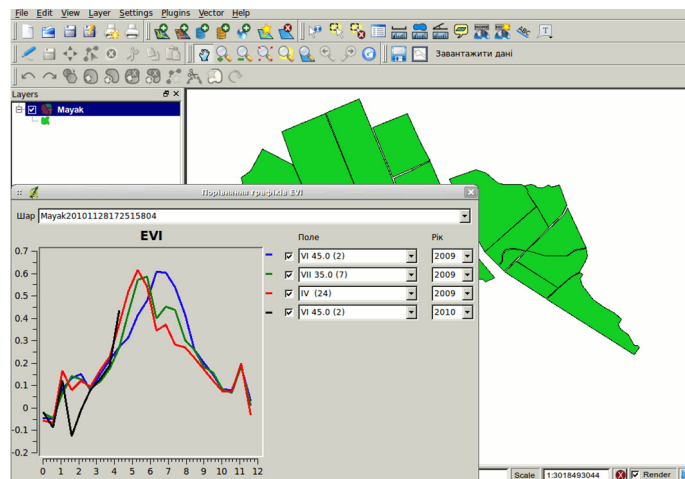


Рис. 6. Сравнение графиков индекса EVI

Еще одним дополнительным компонентом системы является модуль генерирования отчетов. Разработанный дополнительный компонент позволяет осуществлять автоматическое генерирование отчетов в соответствии с шаблоном. Эта возможность может быть использована как для бизнес-аналитики, так и для автоматизации документооборота предприятия (рис. 7).

Вміст поживних речовин

Вміст поживних речовин
(дані з шару fields-2011)

	Name	Descriptio	gumus	N	P	K	Ph	Hr	S	V
0	II	;35.3 ;	0	0	0	0	5.8	1.3	26	95
1	VI	;45 ;	4.94	105	109	6.9	0	0	0	0
2	VI	;154.8 ;	4.94	105	109	6.9	0	0	0	0
3	VI	;27.2 ;	3.76	90	97	66	0	0	0	0
4	VII	;42.3 ;	3.7	85	93	77	0	0	0	0
5	VII	;20.0 ;	3.6	60	81	72	0	0	0	0
6	VII	;35 ;	3.6	60	81	72	0	0	0	0
7	VII	;63.0 ;	2.59	68	63	113	0	0	0	0
8	III	;103.6 ;	0	0	0	0	6.6	0.9	34	97
9	V	/	0	0	0	0	9.6	2	16	89
10	IV	;25.9 ;	0	0	0	0	5.6	0	0	0
11	I	;18.0 ;	0	0	0	0	5.6	0	0	0

Рис. 7. Пример автоматически сгенерированного отчета

Отчеты генерируются в формате HTML и могут содержать информацию о содержании питательных веществ в почве в определенный период времени или количестве внесенных удобрений.

Интерфейс с базой данных. Базовая функциональность QGIS позволяет использовать в качестве источника данных разные форматы, но ее возможности работы со сложно структурированными данными являются ограниченными. Так, поддерживается работа с метаданными, т.е. фиксированным набором полей примитивных типов (число, строка и т.п.), привязанных к одному объекту на карте. Этого недостаточно для работы с общей базой данных распределенной системы, поскольку к каждому полю привязываются сложно структурированные данные: за разный период времени, для разных типов измерений и т.п. Приведение этих данных к формату QGIS повлекло бы потерю структурированности.

Поэтому вместо использования имеющейся в QGIS функциональности для работы с данными, разработан дополнительный модуль, который обеспечивает доступ непосредственно к базе данных SQLite, что сразу же позволяет воспользоваться всеми ее богатыми возможностями, т.е. выполнять

произвольные SQL-запросы. Структура используемой БД приведена и более подробно описана в [Куссуль и др, 2010].

Литература

- [Rembold et al, 2006] F. Rembold, J. Delincé, H. Boogard, A. Burger. Spatial Information Systems in Crop Monitoring: Developing New Global Models and Sharing the Data. GSDI-9 Conference Proceedings. Santiago, Chile, 2006.
- [Kussul et al, 2009] N. Kussul, A. Shelestov, S. Skakun. Grid and sensor web technologies for environmental monitoring. Earth Science Informatics, Vol. 2, N 1-2. P.37-51, 2009.
- [Kussul et al, 2010a] N. Kussul, A. Shelestov, S. Skakun, O. Kravchenko. High performance Intelligent Computations for Environmental and Disaster Monitoring. In Intelligent Data Processing in Global Monitoring for Environment and Security (Krassimir Markov, Vitalii Velychko editors). I T H E A, Sofia, P.64-92, 2010.
- [Kussul et al, 2010b] N.N. Kussul, A.Ju. Shelestov, S.V. Skakun, O.M. Kravchenko, Yu. Gripich, L. Hluchy, P. Kopp, E.A. Loupian. The Data Fusion Grid Infrastructure: Project Objectives and Achievements. Computing and Informatics, Vol. 29. P.319-334, 2010.
- [Popov et al, 2008] M. Popov., N. Kussul, S. Stankevich, A. Kozlova, A. Shelestov, O. Kravchenko, M. Korbakov, S. Skakun. Web Service for Biodiversity Estimation Using Remote Sensing Data. International Journal of Digital Earth, 1(4), P.367-376, 2008.
- [Куссуль и др, 2010] Н.Н. Куссуль, Б.В. Соколов, Я.И. Зельк, В.А. Зеленцов, С.В. Скакун, А.Ю. Шелестов. Оценка рисков стихийных бедствий на основе разнородной геопространственной информации. Проблемы управления и информатики, N 6. С.97-110, 2010.
- [Лупян и др, 2009] Е.А. Лупян, С.А. Барталев, И.Ю. Савин. Технологии спутникового мониторинга в сельском хозяйстве России. Аэрокосмический курьер. № 6. С.47-49, 2009.
- [Толпин и др, 2010] В.А. Толпин, С.А. Барталев, В.Ю. Ефремов, Е.А. Лупян, И.Ю. Савин, Е.В. Флитман. Возможности информационного сервера СДМЗ АПК. Современные проблемы дистанционного зондирования Земли из космоса. Т.7. № 2. С.221-232, 2010.
- [Шелестов и др, 2011] А.Ю. Шелестов, А.Н. Кравченко, С.В. Волошин, и др. Web-портал системы агромониторинга. Наука и инновации. Т. 7, № 3. С. 30-35, 2011.

Информация об авторах

Андрей Шелестов – зав. кафедрой, Национальный университет биоресурсов и природопользования Украины, 03187 Киев, ул. Героев Оборона, 15, e-mail: andrii.shelestov@gmail.com

Наталья Морзе – проректор, Национальный университет биоресурсов и природопользования Украины, 03187 Киев, ул. Героев Оборона, 15, e-mail: nmorze@ukr.net

Ольга Куссуль – аспирант, Национальный технический университет Украины «Киевский политехнический институт», 03056, г.Киев-56, проспект Победы, 37, e-mail: olgakussul@gmail.com

Юлия Грипич – аспирант, Институт космических исследований НАНУ-НКАУ, 03680, г. Киев, проспект Глушкова 40, корп. 4/1, e-mail: yulchenya@ukr.net

ПРОЕКТИРОВАНИЕ БАЗЫ ЗНАНИЙ ДЛЯ СИСТЕМ ОРГАНИЗАЦИОННОГО УПРАВЛЕНИЯ

Баканова Н.Б., Обознов А.А., Баканов А.С.

Абстракт: В представленной работе рассматривается подход, позволяющий создать средства оперативного формирования специализированных баз знаний для организации режимов поддержки принятия решений, на основе данных накопленных в системе организационного управления. В качестве инструмента формирования базы знаний предусматривается использование специальных алгоритмов и средств структурирования информации, извлеченной из баз данных системы.

Ключевые слова: системы организационного управления, базы данных, базы знаний, поддержка принятия решений.

ACM Classification Keywords: I.2.1

Введение

Организационное управление является эффективным инструментом, обеспечивающим стабильное качество выполнения процессов и концентрацию внимания на наиболее важных аспектах деятельности предприятия. Информационные системы, созданные для поддержки управленческих процессов, отличаются масштабами, архитектурой, принципами построения, различной функциональностью. Специфику систем определяет также целевая направленность стратегий управления (бизнес, государственное управление), отраслевые особенности, виды управленческой деятельности и т.д. Необходимость разработки специальных функций, предназначенных для поддержки деятельности конкретной организации, определяет уникальность проектов [Аврамчук Е.Ф. и др., 1988, Вендров А. М., 2003]. Однако, независимо от специфики и вариантов построения системы, в процессе длительного функционирования в базах данных накапливается информация о деятельности организации, об удачных и неудачных решениях, о задачах и поручениях, об исполнителях и качестве исполнения и т.п.

Учитывая важность накопленного материала, практический и научный интерес представляет проблема: проектирования информационных систем, включающих интеллектуальную обработку информационных массивов для реализации функций поддержки принятия управленческих решений, базирующихся на опыте выполнения оперативных задач.

Проблема проектирования информационных систем организационного управления занимает значительное место среди актуальных задач информационных технологий (ИТ). Многогранность организационной деятельности, инвариантность структурного построения, территориальная распределенность, различная целевая направленность, постоянная потребность в совершенствовании и развитии предприятий, породили множество подходов к разработке проектов этих систем. Среди специалистов широко известна классификация информационных систем организационного управления, по типам предоставляемых ими сервисов [Вендров А. М., 2003, Грейди Буч и др., 2000]:

- MRP (Materials Resource Planning) – планирование материальных ресурсов;
- MRP II (Manufacturing Resource Planning) – планирование производственных ресурсов;
- ERP (Enterprise Resource Planning) – планирование ресурсов предприятия;
- CRM (Customer Relationship Management) – система управления отношениями с клиентами;
- SCM (Supply Chain Manament) - система работы с поставщиками;

Новым направлением, в представленной классификации, являются системы класса EDS (Enterprise Development Solution), поддерживающие решения в области организационного развития предприятия. Системы класса EDS должны поддерживать требования стандарта ИСО 9000, регламентирующего характеристики качества и политику организационного управления, а также предусматривать включение в системы интеллектуальных составляющих, для совершенствования процессов управленческой деятельности, определения и реализации задач развития.

В данной статье рассматривается проблема, которая относится к задачам систем класса EDS, и посвящена разработке метода формирования базы знаний, на основе данных накопленных системой организационного управления, а также разработке метода представления этого знания в режиме поддержки принятия решений.

Извлечение знаний на основе данных об управленческих процессах

Концептуально задача создания такого проекта включает вопросы интеграции систем двух типов: системы организационного управления и системы поддержки принятия решений. В проекте должны быть предусмотрены; разработка принципов извлечения необходимых данных из оперативных баз, формирование базы знаний по выбранному направлению, разработку сервисов, обеспечивающих интеллектуальные интерфейсы работы. Результат решения позволит повысить эффективность интегрированной системы за счет использования интеллектуальных функций, обеспечит сервисы при принятии управленческих решений.

Программный комплекс систем организационного управления, по назначению и функциям можно разделить на две группы подсистем. Первая группа – оперативные подсистемы - предназначена для решения функциональных задач управленческих подразделений и общих – информационных задач организации. Вторая группа – административно-организационные подсистемы - предназначена для управления информационным комплексом, синхронизации оперативных подсистем, анализа и обобщения данных, обеспечения жизненного цикла программного комплекса. Количество, масштаб и конкретное назначение подсистем каждой группы зависит от целей и задач объекта управления [Bakarova N., Atanasova T., 2008].

Назначением оперативных подсистем является автоматизация различных управленческих процедур, информационная и технологическая поддержка, выполняемых в организации работ, накопление информации о процессах выполнения и результатах различных управленческих процессов. Среди основных функциональных модулей, традиционно включаемых в организационные системы, можно выделить следующие подсистемы:

- организация информационных процессов прохождения документов;
- организация поддержки финансовой деятельности;
- управление кадровыми службами;
- управление производственной деятельностью;
- задачи планирования различного целевого направления;
- и другие.

В результате функционирования системы в базе данных накапливается информация (описательная, количественная, качественная) о процессах и особенностях выполнения конкретных управленческих заданий, об опыте решения задач, об участниках процессов. Извлечение знаний будет базироваться преимущественно на информации, накопленной основными функциональными модулями.

Информация, накопленная в базах данных системы, представлена в виде различных текстовых фрагментов, например: краткое содержание, резолюции, ссылки на другие документы, данные по процессам исполнения руководящих решений, анкеты т.д. Совокупность данных отражает деятельность

организации за достаточно продолжительный период времени. Результаты анализа этих данных могут быть полезны для решения различных задач организации, в том числе и для выявления динамики основных направлений деятельности.

Для исследования содержимого баз данных, предлагается метод целенаправленной обработки текстовых фрагментов документов, позволяющий определить связь документа с одним или несколькими заданными направлениями деятельности, а также дать оценку отношения документа к тому или иному направлению. Предлагаемый метод решения базируется на идее использования и согласования ряда известных подходов к исследованию текстовых массивов, а также включает разработку показателей оценки полученных результатов.

Тематические направления деятельности организации задаются в виде некоторого классификатора, формирование которого является частью решения задачи. При разработке метода в качестве классификатора был выбран проблемно-ориентированный тезаурус, который создается на основе тщательного анализа и разработки тематических направлений, наиболее важных в деятельности данной организации.

Тезаурус представляет собой совокупность словарей нормативной лексики, термины которых связаны сетевой структурой, то есть, помимо родовидовых отношений, в тезаурусе используются ассоциативные связи. Графически тезаурус можно представить в виде ориентированного графа, в узлах которого находятся термины некоторой предметной области, а ребра могут иметь различные названия в зависимости от направления связи [Баканова Н.Б., Шелов С.Д., 2010]. В задачи тезауруса информационной системы входит возможность использования отношений между терминами для реализации поисковых стратегий (расширение или сужение области поиска). Это свойство тезауруса будет полезно при анализе совпадения терминов тезауруса и терминов текстовых фрагментов базы данных. Правила разработки, структура, состав и форма представления одноязычного информационно-поискового тезауруса регламентируются ГОСТ 7.25-2001.

В процессе обработки текстовых фрагментов требуется получение оценок отношения найденной информации к тому или иному тематическому направлению. Разработка оценочных показателей является одной из ответственных задач данного метода. В организационном управлении данные часто являются неточными, не имеют очевидных количественных оценок.

Задача получения знаний из информации, накопленной в системе, базируется на использовании методов и алгоритмов обработки данных, применяемых в системах поддержки принятия решений и адаптированных к условиям работы систем организационного управления [Ларичев О.И., 2002, Люггер Д.Ф., 2003].

В частности, при обработке текстовых фрагментов документов оценочные характеристики получают с помощью нечетких величин на основе комбинаций „IF-THEN” правил, которые устанавливаются на базе экспертных процедур. Например, при разработке алгоритма оценки документопотока организации выбраны следующие количественные (кол.) и качественные (кач.) показатели, которые можно получить при анализе документов:

- соответствие документа выбранным тематическим направлениям (кол./кач.);
- пересечение документа с другими тематическими направлениями (кол./кач.);
- авторитет автора документа или коллектива авторов (кач.);
- наличие восходящих связей документа с другими документами базы данных (кол./кач.);
- наличие нисходящих связей документа с другими документами базы данных (кол./кач.);
- наличие уровней контроля (кол./кач.);
- важность документа для организации (кач.);
- и другие.

Модуль оценки реализует использование правил нечеткой логики. Система качественных показателей основывается на лингвистических (описательных) переменных. Описательные переменные представлены шкалами нечетких величин, которые оцениваются экспертными процедурами. Совокупность правил определяет ценность документа по каждому тематическому направлению. Ниже показаны примеры формирования комбинаций „IF-THEN” правил для анализа документопотока.

Правило 1:

IF документ использует термины тезауруса

AND (частота использования терминов высока

OR частота встречаемости терминов других тематических направлений высока

OR документ поступил из вышестоящих организаций)

THEN оценка документа высока.

Правило 2:

IF документ использует термины тезауруса

AND (количество восходящих связей документа высоко

OR количество нисходящих связей документа высоко

OR документ находится на контроле высокого уровня)

THEN оценка документа высока.

и т.д.

Системы поддержки принятия решений, как правило, используют базы знаний, формирование которых идет под контролем специалистов предметной области, что требует определенных ресурсов: времени для обработки данных, наличия экспертов, времени на реструктуризацию базы знаний и т.д. Предлагаемый алгоритм позволяет формировать базу знаний в процессе поступления и обработки информации в системе организационного управления.

Представление знаний в системах организационного управления

При проектировании базы знаний в системах организационного управления, внимание должно быть уделено вопросам представления знаний, а также использования знаний при поиске решения конкретной проблемы [Емельянов С. В., Ларичев О. И., 1985, Ларичев О. И., Петровский А. Б., 1987]. Использование разных способов представления имеющихся знаний, как правило, является существенным плюсом программной реализации базы знаний. Однако только в исключительно редких случаях, бывает так, чтобы конкретный способ решения проблемы можно было одинаково хорошо представить всеми возможными способами. Поэтому выбор способа представления и стратегии представления оказывает существенное влияние на процесс принятия решения [Обознов А.А., Баканов А.С., 2010]. Стратегии представления определяют, каким способом, и в какой последовательности программная реализация базы знаний будет представлять варианты решения в некотором пространстве альтернатив. Для представления знаний в системах организационного управления, целесообразно использовать следующие варианты/методы (или их комбинации):

- Представление знаний в виде семантической сети;
- Представление знаний в виде фреймов;
- Представление знаний в виде последовательности событий/действий (скрипт).

Наиболее подходящим для решения поставленной задачи был признан метод представления знаний в виде семантической сети. Каждый узел такой сети представляет вид деятельности (процесс), а ребра определяют отношения между процессами. Использование метода семантической сети позволяет

наглядно представить, как данный способ решения проблемы связан с альтернативными способами, давая возможность (посредством пользовательского интерфейса) динамически перемещаться между вершинами графа и оценивать возможный результат. Необходимо отметить, что пользовательский интерфейс, (каким бы он ни был дружественным и интуитивно понятным) не решает проблем структуризации и корректности процедур получения информации от пользователя.

Выводы

Идея интеграции прикладных информационных систем организационного управления с модулями «интеллектуальной поддержки управления», которой посвящена данная статья, базируется на главной функции систем организационного управления, заключающейся в накоплении взаимосвязанной информации об управленческих процессах, проходящих в организации. Реализация идеи позволит расширить возможность использования накопленного информационного потенциала, существенно увеличит функциональность системы за счет интеллектуальных сервисов поддержки принятия решений (ППР).

Благодарности

Работа опубликована при финансовой поддержке проекта ITHEA XXI Института информационных теории и приложений FOI ITHEA (www.ithea.org) и Ассоциации разработчиков и пользователей интеллектуальных систем ADUIS Украины(www.aduis.com.ua).

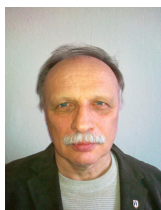
Литература

- [Аврамчук Е.Ф. и др., 1988] Аврамчук Е.Ф., Вавилов А.А., Емельянов С.В. и др. Технология системного моделирования М., 1988
- [Bakanova N.,Atanasova T., 2008] Bakanova N.,Atanasova T. Development of the combinat metod for dataflow system. IJ "Information technologies & knowledge". Volume 2, № 3, p.262-266, 2008
- [Вендров А. М., 2003] Вендров А. М. Проектирование программного обеспечения экономических информационных систем., М.: Финансы и статистика, 2003.
- [Грейди Буч и др., 2000] Грейди Буч, Джеймс Рамбо, Айвар Джекобсон. Язык UML. Руководство пользователя., М., ДМК, 2000. (с. 429)
- [Ларичев О.И., 2002] Ларичев О.И. Теория и методы принятия решений. 2-е изд. М.: Логос, 2002. – 392 с.
- [Люггер Д.Ф., 2003] Люггер Д.Ф. Искусственный интеллект. Стратегия и методы решения сложных проблем. Пер. с англ. М.: Вильямс, 2003.
- [Баканова Н.Б., Шелов С.Д., 2010] Баканова Н.Б., Шелов С.Д. Использование контент-анализа для исследования результатов работ по инвестиционным проектам, Материалы Четвертой международной конференции «Управление развитием крупномасштабных систем MLSD'2010», Москва, ИГУ РАН, 2010, стр. 206-209.
- [ГОСТ 7.25-2001] ГОСТ 7.25-2001. Межгосударственный стандарт «Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления».
- [Емельянов С. В., Ларичев О. И., 1985] Емельянов С. В., Ларичев О. И. Многокритериальные методы принятия решений. – М., Знание. 1985г.,
- [Ларичев О. И., Петровский А. Б., 1987] Ларичев О. И., Петровский А. Б. Системы поддержки принятия решений современное состояние и перспективы развития. // Техническая кибернетика т. 21. – М., ВИНТИ. 1987г.,
- [Обознов А.А., Баканов А.С., 2010] Обознов А.А., Баканов А.С. Разработка экспертной системы на основе концептуальной модели знаний у операторов атомных электростанций // Труды конгресса по интеллектуальным системам и информационным технологиям том 1.: – М., Физматлит, 2010. с.290-294.

Сведения об авторах



Нина Борисовна Баканова – к.т.н., доцент, Институт проблем передачи информации РАН, Б. Каретный пер., Москва, 127994, Россия; e-mail: nina@iitp.ru
Major Fields of Scientific Research: General theoretical information research, Software technologies,



Александр Александрович Обзнов – д.п.н., Институт психологии РАН, Ярославская ул., Москва, 129366, Россия; e-mail: aao46@mail.ru
Major Fields of Scientific Research: Business informatics, Ergonomics



Арсений Сергеевич Баканов – к.т.н., Институт психологии РАН, Ярославская ул., Москва, 129366, Россия; e-mail: ars@iitp.ru
Major Fields of Scientific Research: Intelligent Systems, Software technologies

МОДЕЛЬ ИЗВЛЕЧЕНИЯ ЗНАНИЙ ИЗ НЕСТРУКТУРИРОВАННЫХ ДОКУМЕНТОВ КОРПОРАТИВНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Нина Хайрова, Наталья Шаронова

Аннотация: В работе предлагается математическая модель отношений между областью интеллектуальной деятельности менеджера и достоверными глубинными знаниями, представленными в документах, поступающих в корпоративную информационную систему (КИС). Доказывается возможность использования математического аппарата алгебры конечных предикатов в качестве базового средства описания модели. В работе показана возможность использования данной модели для извлечения знаний, представленных терминологическими понятиями и отношениями между ними, из разноформатных текстов документов КИС, с их одновременным динамическим разбиением на персонафицированные локальные области знаний менеджеров корпорации. Описывается информационное и алгоритмическое обеспечение программной реализации модели, приводится расчет эффективности использования разработанной подсистемы.

Ключевые слова: алгебра конечных предикатов, извлечение знаний, динамическая классификация, неструктурированная информация, корпоративная информационная система.

ACM Classification Keywords : H.3.3 .Information Search and Retrieval

Введение

В последние годы достигнуты большие успехи в развитии КИС, основным направлением повышения эффективности работы которых, становится разработка интеллектуальных моделей трансформации информации, доступной снаружи и внутри организации, в инновационные знания. На передний план выходят требования по добыче «новых» для данной отрасли знаний, обеспечивающих фирму конкурентным потенциалом для принятия управленческих решений в изменяющемся бизнес-сообществе. Такого рода знания, в большинстве своем, представлены в текущих неструктурированных или слабоструктурированных документах организации. При этом, несмотря на достижения Text Mining, задача извлечения знаний из неструктурированной текстовой информации по-прежнему остается актуальной.

Современное понимание КИС подразумевает совокупность различных программно-аппаратных платформ, универсальных и специализированных приложений, интегрированных в единую информационно однородную среду, позволяющую принимать оптимальные для конкретного предприятия управленческие решения в соответствии с формализованными методами и правилами менеджмента. Одним из основных направлений повышения эффективности работы корпоративной информационной системы сегодня становится разработка интеллектуальных моделей трансформации информации, поступающей в организацию, в инновационные знания, представляющие интеллектуальные активы компании.

Рассматривая современные КИС с точки зрения управления знаниями можно заметить, что основной акцент сегодня делается не на сохранении разрозненной информации, а на извлечении закономерностей и принципов, позволяющих решать производственные и бизнес задачи, т.е. осуществлять накопление знаний [Гаврилова, 2000], [Mancini, 2001]. Источником знаний, как правило, служат документы, поступающие в систему на обработку: корпоративные стандарты, методики, бизнес-правила и технологии, а также технологическая и трудовая документация, накопившаяся в процессе функционирования организации. При этом перед КИС ставится задача извлечь и накопить именно инновационные знания,

которые снабжают фирму конкурентным потенциалом, а не коренные, устоявшиеся и даже «старые» знания, которыми обладают все участники данной отрасли. Чаще всего подобные «новые» знания, в большинстве своем, заложены в текущих, ежедневно поступающих в организацию документах.

Таким образом, основной задачей повышения эффективности КИС становится разработка системы трансформации информации, доступной снаружи и внутри организации в интеллектуальные архивы компании, представляющие инновационные знания. И если для решения данной задачи при работе со структурированной информацией используются хорошо разработанные технологии и методы data mining, то существующие сегодня методы извлечения знаний из неструктурированных полнотекстовых документов не позволяют достаточно качественно разрабатывать единое информационное пространство, представляющее модель знаний предметной области работы КИС [Carroll, 2005].

Описание математической модели

В качестве базовых средств модели используем средства алгебры конечных предикатов (АКП). Вводим универсум элементов U , включающий все возможные текстовые документы, поступающие в корпоративную информационную систему на обработку, а также понятия и объекты анализа рассматриваемой предметной области, специализированные словари, тезаурусы, отображающие специфику данной предметной области. С данным универсумом U связаны некоторые знания, относящиеся к персонифицированному интеллектуальному ресурсу менеджера, т.к. они включены во все возможные текстовые документы, поступающие на обработку (справки, выписки, отчеты, распоряжения, решения и т.д.).

Из элементов универсума в соответствии с конкретной задачей обработки информации образуются подмножества $M_{1i}, M_{2i}, \dots, M_{mi}$, на декартовых произведениях которых $M_{1i} \times M_{2i} \times \dots \times M_{mi}$ определяются предикаты P_j , характеризующие работу системы.

Предикатом P , заданным на универсуме, будем называть любую функцию, отображающую множество элементов универсума в ноль (предикат тождественно ложный) или единицу (предикат тождественно истинный). Так как множество элементов универсума информационной системы корпорации является конечным, то и предикат P соответственно конечен.

Базисным для алгебры предикатов является предикат узнавания предмета a по переменной x_i , равный единице, в том случае, если x_i равен a и нулю в противном случае, где i — это любой элемент универсума [Бондаренко, 2007]:

$$x_i^a = \begin{cases} 1, & \text{если } x_i = a \\ 0, & \text{если } x_i \neq a \end{cases}, \quad (1 \leq i \leq n) \quad (1)$$

Алгебра предикатов полна в том смысле, что любой ее предикат можно представить в виде суперпозиции базисных операций, примененных к базисным элементам. В работе [Шабанов-Кушнарченко, 1986] показано, что на языке АКП могут быть описаны любые конечные отношения, поэтому любой другой математический аппарат, предназначенный для описания произвольных конечных отношений, в логическом смысле, обязательно будет эквивалентен алгебре конечных предикатов.

Для построения модели описания персонифицированной локальной области знаний менеджера вводится конечное множество документов $D = \{d_i\}$, $1 \leq i \leq n$, поступающих менеджеру на обработку, и конечное множество терминологических понятий $T^k = \{t^k_j\}$, $1 \leq j \leq m$, первоначально определяющих область деятельности k -го менеджера, и представляющих его интеллектуальный ресурс, который был апробирован, разработан и использован данным или вышестоящим менеджером.

Под документом понимается некоторая целостная единица информации, имеющая уникальный идентификатор, средства отображения и модификации. Под терминологическим понятием понимается совокупность суждений о каком-либо объекте, отражающая его сущность, и выделяющая предметы некоторого класса по общим и совокупным специфичным для данного класса признакам.

На декартовом произведении $D \times T$ множеств D и T^k вводится бинарный предикат персонификации интеллектуального ресурса компании $P(d, t)$, однозначно задающий отношение между обрабатываемым документом и терминологическим понятием интеллектуального ресурса менеджера:

$$P(d_i, t_j^k) = \delta \mid d_i \in D, t_j^k \in T^k, \delta = \{0, 1\}. \quad (2)$$

Классификатор последовательно анализирует все возможные пары из множества $D \times T^k$, однозначно приравнивая любой предикат $P(d_i, t_j^k)$ к нулю или к единице. Значение δ является истинным, если терминологическое понятие рассматриваемого документа определяет персонифицированную область деятельности менеджера, и является ложным в противном случае. Рассматриваемый предикат P удовлетворяет постулату существования [Хайрова, 2003]: предикат $P(d_i, t_j^k)$ реально существует в том и только в том случае, если при повторной итерации любой пары (d, t) из множества $D \times T^k$ классификатор будет реагировать тем же ответом, что и в первый раз.

Используя результаты исследования [Шабанов-Кушнарченко, 1993], для предиката персонификации интеллектуального ресурса компании можно ввести два определения. Два документа d_v и d_w относятся к области знаний k -го менеджера $(d_v, d_w) \in D$ тогда и только тогда, когда для $\forall t^k: P(d_v, t^k) = P(d_w, t^k)$. В этом случае можно говорить о том, что два документа тождественны по отношению к области знаний менеджера, и писать $d_v \sim d_w$. Два терминологических понятия t_v^k и t_w^k относятся к области интеллектуального ресурса k -го менеджера $(t_v^k, t_w^k) \in T^k$ тогда и только тогда, когда для $\forall d: P(d, t_v^k) = P(d, t_w^k)$. В этом случае можно говорить о том, что два терминологических понятия тождественны по отношению к персонифицированному интеллектуальному ресурсу k -го менеджера, и писать $t_v^k \sim t_w^k$.

Рассмотрев все возможные пары на декартовом произведении $d \times t^k$, получаем отображение Ω , множества рассматриваемых документов D на множество терминологических понятий T^k персонифицированного интеллектуального ресурса менеджера. Отображение Ω можно представить в виде двудольного графа, верхнее множество вершин которого представляет множество рассматриваемых документов D , нижнее множество вершин — множество терминологических понятий T^k интеллектуального ресурса k -го менеджера, а дуги показывают истинность предиката персонифицированного интеллектуального ресурса компании $P(d_i, t_j^k) = 1$. На декартовом квадрате $D \times D$ универсума U вводим предикат соответствия документов персонифицированному интеллектуальному ресурсу k -го менеджера:

$$G_1(d_1, d_2) = \forall t^k \in T^k (P(d_1, t^k) \sim P(d_2, t^k)). \quad (3)$$

А на декартовом квадрате $T^k \times T^k$ универсума U вводим предикат соответствия терминологических понятий персонифицированному интеллектуальному ресурсу k -го менеджера:

$$G_2(t_1^k, t_2^k) = \forall d \in D (P(d, t_1^k) \sim P(d, t_2^k)). \quad (4)$$

Предикаты G_1 и G_2 , определяемые выражениями (3) и (4) однозначно определяются предикатом P , рефлексивны, транзитивны и симметричны, из чего следует, что они являются предикатами эквивалентности. Предикат (3) можно использовать для объективного определения отношения двух любых документов d_1 и d_2 , принадлежащих множеству D , к одной области знаний менеджера. Если $G_1(d_1, d_2)=1$, то при любом терминологическом понятии $t^k \in T^k$: $P(d_1, t^k) = P(d_2, t^k)$, т.е. информация данных документов, выражаемая терминологическими понятиями из множества T^k , позволяет отнести их к области знаний k -го менеджера. В противном случае, если $G_1(d_1, d_2)=0$, то найдется такое терминологическое понятие $t^k \in T^k$, для которого $P(d_1, t^k) \neq P(d_2, t^k)$. В этом случае информация, передаваемая документами $d_1, d_2 \in D$, выражаемая терминологическими понятиями t^k множества T^k , не совпадает и, следовательно, документы нельзя отнести к области знаний одного менеджера.

Предикат (4) можно использовать для определения соответствия двух любых терминологических понятий, принадлежащих множеству T^k , персонифицированному интеллектуальному ресурсу одного менеджера. Действительно, если $G_2(t^k_1, t^k_2)=1$, то $P(d, t^k_1) = P(d, t^k_2)$ для любого документа $d \in D$, то есть существуют терминологические понятия t^k_1 и t^k_2 , которые либо одновременно относятся к области знаний k -го менеджера, либо одновременно не относятся. Иначе говоря, во множестве документов D нет такого документа $d \in D$, который бы одновременно и относился и не относился к области знаний k -го менеджера, выражаемой терминологическими понятиями $t^k \in T^k$. Если же $G_2(t^k_1, t^k_2)=0$, то найдется такой документ $d \in D$, для которого $P(d, t^k_1) \neq P(d, t^k_2)$. То есть либо документ d включает знания, выражаемые терминологическим понятием t^k_1 и не включает знания, выражаемые терминологическим понятием t^k_2 , либо, наоборот, документ d , включает знания, выражаемые терминологическим понятием t^k_2 и не включает знания, выражаемые терминологическим понятием t^k_1 . В обоих случаях терминологические понятия t^k_1 и t^k_2 будут относиться к персонифицированному интеллектуальному ресурсу разных менеджеров.

Очевидно, что документы, входящие в полученные классы эквивалентности, не могут быть тождественными по смыслу безотносительно от терминологических понятий, отображающих область знаний того или иного менеджера, они являются эквивалентными относительно персонифицированной области знаний менеджера. Используя предикаты (3, 4), можно определить разбиение множества D на слои документов, а множества T — на слои терминологических понятий персонифицированных интеллектуальных ресурсов менеджеров корпорации. Класс \mathcal{G}_c всех документов $d \in D$, относящихся к области знаний одного менеджера, включающий документ $c \in D$, можно выразить как:

$$\mathcal{G}_c(d) = \forall t \in T (P(d, t) \sim P(c, t)). \quad (5)$$

Класс Λ_b всех терминологических понятий $t^k \in T$, относящихся к персонифицированному интеллектуальному ресурсу k -го менеджера, включающий терминологическое понятие $b \in T$, можно представить следующим образом:

$$\Lambda_b(t) = \forall d \in D (P(d, t) \sim P(d, b)). \quad (6)$$

Формулы (5, 6) выражают разбиение документов, поступающих на обработку, и терминологических понятий, относящихся к области интеллектуальных активов менеджеров корпораций, на классы эквивалентности через предикат персонифицированного интеллектуального ресурса компании (2), однозначно определяемый классификатором.

Пример реализации модели

В рассматриваемой модели предметными переменными, формирующими терминологические понятия и определяющими отношение документа к предметной области деятельности менеджера, являются: l — ключевое слово или словосочетание документа; u — значение универсальной иерархической классификации (УДК) и r — предметная рубрика рубрикатора, принятого для систематизации областей знаний потока научно-технической информации. Поскольку данные переменные отражают суть документа, назначение и взаимосвязь его составляющих, можно говорить о том, что они объективно определяют истинные и достоверные глубинные знания, представлены в документах.

Также в модели используется, основное для наших рассуждений, понятие области знаний менеджера q . Под областью интеллектуальных знаний конкретного менеджера организации будем понимать нечетко определенную часть корпоративных знаний, используемую для решения стандартных управленческих задач данного менеджера. Как известно, современная КИС является многопользовательской, то есть одновременно используемой конечным множеством топ-менеджеров данной организации при решении различных управленческих задач, при этом области деятельности менеджеров далеко не всегда совпадают с принятыми предметными областями. Под областью интеллектуальных знаний менеджера будем понимать совокупность некоторых предметных областей знаний, являющихся информационным обеспечением управленческих задач конкретного менеджера и выделяемых в некоторый класс по определенным общим и совокупным специфичным для данного менеджера управленческим ситуациям.

Например, при описании информационного пространства десяти документов, относящихся к финансово-экономической деятельности предприятия, на этапах предлингвистического анализа статистико-позиционными методами [Khairova, 2009] определяются: множество ключевых слов $L=\{l^i\}$, $1 \leq i \leq 14$, где l^1 = депозитные операции, l^2 = депонирование, l^3 = банк, l^4 = вексель,..., l^{14} = цензовая стоимость; множество значений $U=\{u^i\}$, $1 \leq i \leq 5$, где $u^1=336.22$ Налоги. Сборы, $u^2= 336.24$ Таможенные пошлины, $u^3= 336.71$ Банковское дело. Банки, $u^4 =336.763$ Ценные бумаги. Активы, $u^5= 336.764$ Биржевые операции; множество значений рубрикатора $R =\{r^i\}$, $1 \leq i \leq 4$, где r^1 = банки и банковская деятельность; r^2 =рынок ценных бумаг; r^3 = экономика фирмы; r^4 = финансовая экономика. Множество областей интеллектуальных знаний менеджера, обусловленное множеством рассматриваемых управленческих ситуаций, $Q=\{q^i\}$, $1 \leq i \leq 18$, где q^1 = банковские операции, q^2 = транзитная торговля, q^3 = депонирование, ..., q^{18} = налогообложение.

Можно построить парадигматическую таблицу, отображающую связь между областью локализации деятельности менеджера q^i и предметными переменными l , u и r . Используя данную таблицу отношений, выразим область локализации интеллектуальной деятельности менеджера q через значения предметных переменных r , l и u :

$$\begin{aligned} r^1 u^1 l^{12} = q^1; r^4 u^2 l^{12} = q^2; r^1 u^1 l^{13} = q^3; r^1 u^3 l^{14} = q^4; r^1 u^3 l^1 = q^5; r^2 u^4 l^1 = q^6; r^1 u^3 l^2 = q^7; r^1 u^3 l^3 = q^8; r^1 u^3 l^4 = q^9; \\ r^2 u^4 l^4 = q^{10}; r^2 u^4 l^5 = q^{11}; r^2 u^4 l^6 = q^{12}; r^2 u^5 l^7 = q^{13}; r^2 u^5 l^8 = q^{14}; r^1 u^3 l^8 = q^{15}; r^4 u^2 l^9 = q^{16}; r^4 u^2 l^{10} = q^{17}; r^4 u^2 l^{11} = q^{18}. \end{aligned} \quad (7)$$

Выполняем операцию почленной дизъюнкции как можно большего количества родственных равенств. Родственными равенствами будем называть такие равенства, из которых после выполнения над ними операции почленной дизъюнкции можно получить равенства с левой частью в виде логического произведения, каждый сомножитель которого зависит от одной предметной переменной [Бондаренко, 2007]. Введение почленной дизъюнкции с использованием родственных равенств обусловлено необходимостью получения локальных областей интеллектуальных знаний менеджера, определяемых именем конкретного менеджера. Такие области могут включать больше чем одно исчисляемое ограниченное количество рубрик и предметных областей исследований.

$$\begin{aligned}
 r^1 u^1 (l^{12} \vee l^{13} \vee l^{14}) &= q^1 \vee q^3 \vee q^4; & r^4 u^2 (l^{12} \vee l^9 \vee l^{10} \vee l^{11}) &= q^2 \vee q^{16} \vee q^{17} \vee q^{18}; & r^2 u^5 (l^7 \vee l^8) &= q^{13} \vee q^{14} \\
 r^1 u^3 (l^1 \vee l^2 \vee l^3 \vee l^4 \vee l^8) &= q^5 \vee q^7 \vee q^8 \vee q^9 \vee q^{15}; & r^2 u^4 (l^1 \vee l^4 \vee l^5 \vee l^6) &= q^6 \vee q^{10} \vee q^{11} \vee q^{12}.
 \end{aligned}
 \tag{8}$$

Формируя функцию перехода от предметной области интеллектуальных знаний q к локальной области исследования менеджера m , в профессиональную деятельность которого входит данная область исследования q , и переопределяя зависимость локальной области исследования менеджера m от переменных r, l, u , получаем предикат $P(r, l, u, m)$, описывающий связь локальной области исследования менеджера и переменных, объективно определяющих глубинные знания документа:

$$\begin{aligned}
 P(r, l, u, m) &= m^1 r^1 u^1 (l^{12} \vee l^{13} \vee l^{14}) \vee m^1 r^2 u^5 (l^7 \vee l^8) \vee m^2 r^4 u^2 (l^{12} \vee l^9 \vee l^{10} \vee l^{11}) \vee \\
 &\vee m^3 r^1 u^3 (l^1 \vee l^2 \vee l^3 \vee l^4 \vee l^8) \vee m^4 r^2 u^4 (l^1 \vee l^4 \vee l^5 \vee l^6).
 \end{aligned}
 \tag{9}$$

Данный предикат наглядно изображается в виде логической сети (рис. 1), которая является графической интерпретацией результата бинарной декомпозиции многоместного предиката.

Каждому полюсу логической сети ставится в соответствие своя предметная переменная модели. С каждым полюсом связана область изменения атрибута этого полюса. Любой полюс логической сети в каждый момент времени несет определенное знание о значении своего атрибута. Каждой ветви логической сети ставится в соответствие свое бинарное отношение модели, которое называется отношением этой ветви. Каждая ветвь соединяет два полюса, отвечающие тем предметным переменным, которые связываются отношением, соответствующим данной ветви.

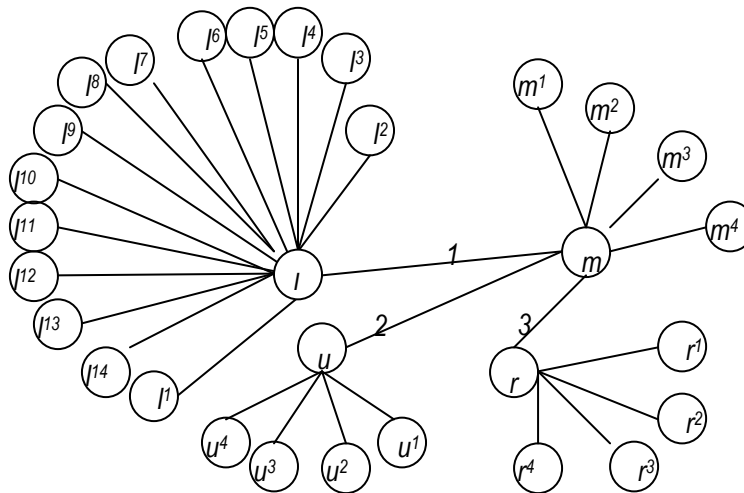


Рис. 1. Логическая сеть формального описания локальной области деятельности менеджера

Программная реализация модели

Предложенный метод разбиения предметной области корпоративной информационной системы на локальные области деятельности менеджера реализуется в программном комплексе, включающем семь логических этапов, и представляющем семантически ориентированный лингвистический процессор: предлингвистическая обработка; графемная обработка; морфологическая обработка; контекстный анализ;

статистическая обработка; логико-алгебраическая обработка обучающей выборки; динамическая классификация поступающей информации.

На этапе графемной обработки выделяются элементы текста, имеющие графемное значение: заголовок, подзаголовок, абзац, предложение и т.д. Блок морфологической обработки вводится для учета словоизменительных форм и представления слов в канонической форме. Морфологический анализ осуществляется методом квазиокончаний. На этапе контекстного анализа из множества лексем выбираются словосочетания. Для учета информационной значимости ключевых слов на этапе статистической обработки вводятся весовые коэффициенты, являющиеся дополнительным средством семантической дифференциации лексических единиц документа. На этапе логико-алгебраической обработки обучающей выборки для описания связи предметной области деятельности менеджера, работающего с документами, с предметными переменными, объективно определяющих глубинные знания документа, используется полученное на предыдущих этапах работы системы информационное представление каждого документа в виде множества ключевых слов и словосочетаний, значений УДК и рубрик.

Строятся бинарные предикаты: $P_l(l,m)$, $P_r(r,m)$, и $P_u(u,m)$. Предикаты P_l , P_r и P_u можно представить в виде таблицы, заполняемой единицами или нулями в зависимости от значений соответствующих предикатов при данных значениях предметных переменных l , u , r , m . На практике часто встречаются ситуации, когда, исключая из рассмотрения некоторые элементы декартового произведения, можно получить разбиения множеств, более соответствующие интуитивным представлениям специалистов о семантике ключевых слов, предметных рубрик и значений УДК. Исключение некоторых элементов декартова произведения модели имеет смысл в том случае, когда упорядоченных пар не много по сравнению с общим числом элементов декартового произведения. Разработанный метод построения разбиений множеств L , R и U учитывает такие исключения. Алгоритм допускает небольшие различия между строками таблиц, попадающими в один класс. Мера таких допустимых отклонений ρ' может устанавливаться пользователем. При этом те строки (или столбцы), которые могут быть отнесены к различным классам, выделяются особо и могут быть классифицированы пользователем отдельно. Такие строки (столбцы) отличаются от элементов некоторых классов на ρ двоичных разрядов, не превышающем заданного ρ' ($\rho \leq \rho'$), где ρ можно интерпретировать как расстояние между векторами: $\rho(a,b) = \sum \alpha_i \oplus \beta_i$; a и b — двоичные векторы: $a = (\alpha_1, \alpha_2, \dots, \alpha_k)$, $b = (\beta_1, \beta_2, \dots, \beta_k)$.

На последнем шаге этапа обучения полученные классы ключевых слов, УДК и значений рубрикатора вводятся в эталоны, представляющие множества U_1, U_2, \dots, U_n , относящиеся к областям исследования конкретных менеджеров. Разбивая таким образом терминологические понятия обрабатываемых КИС документов, мы иерархически упорядочиваем их, используя этого отношения «быть элементом класса» и «включаться в предметную область исследования».

Для последующей динамической классификации множества документов, поступающих на вход системы, выполняются процедуры предлингвистической, графемной, морфологической, контекстной и статистической обработки, в результате которых каждому документу приписывается его информационное представление. Этап сравнения полученного информационного представления с имеющимися эталонами позволяет отнести документ к соответствующим областям деятельности того или иного менеджера. Для вновь полученных документов можно повторить этап обучения для создания обновленных эталонов множеств терминологических понятий U_1, U_2, \dots, U_n , относящихся к областям исследования конкретных менеджеров. При этом система может гибко и динамично менять разбиение понятий и объектов, обрабатываемых менеджером в процессе решения управленческих задач, и, соответственно, менять систему классификации документов, поступающих на вход КИС.

Так как в настоящее время нет четких и однозначных определений понятий «эффективного извлечения знаний» и «качества знаний», количественная оценка результатов работы системы не является тривиальной задачей. Для оценки эффективности работы системы используем метод тестовых коллекций [Сормаск, 1998], заключающийся в сравнении выводов, сделанных системой, с мнением экспертов.

Экспертами, которыми являлись менеджеры корпорации, работающие с документами в КИС, определяется релевантность отнесения документов к области деятельности менеджера. Отношение релевантности является субъективным сложно определяемым, при определении эффективности работы системы, мы основывались на определении релевантности, используемом в [Mizzaro, 1996], согласно которому релевантность зависит от четырех понятий: Relevance (IR, IN, C, T), где IR — информационный ресурс, IN — информационная потребности, C — контекст и T — время. В предлагаемой модели информационный ресурс представлен множеством электронных документов, поступившим на обработку $IR = D$. Для получения интегральных показателей качества работы системы перевода информации в интеллектуальный активы компании применялись методики усредненных метрик [Кураленок, 2002]. В качестве метрик использовались коэффициенты точности, полноты и шума, а для вычисления среднего гармоничного точности и полноты использовалась мера Ван Ризбергена.

Исследовалась выборка из трехсот документов, поступающих в организацию, в среднем в течение десяти дней. Средний коэффициент полноты, определяемый отношением числа релевантных документов, отнесенных к области деятельности менеджера, к общему числу поступивших в систему документов, релевантных области деятельности данного менеджера, $recall = 0,91$. Средний коэффициент точности, определяемый отношением числа релевантных документов, отнесенных к области деятельности данного менеджера, к общему числу документов, отнесенных к области деятельности данного менеджера, $precision = 0,825$. Коэффициент шума, определяемый отношением числа нерелевантных документов, отнесенных к области деятельности менеджера, к общему числу документов, отнесенных к области деятельности данного менеджера, $error = 0,03$. Сбалансированная мера Ван Ризбергена была определена как $F_1 = 0,8654$.

Сравнивая полученные показатели с коэффициентами полноты и точности аналогичных систем, следует отметить, что если значения коэффициента полноты современных гипертекстовых информационно-поисковых систем в среднем $recall \geq 0,9$, а точности $precision \geq 0,8$, то значения данных коэффициентов для документальных систем значительно ниже, в среднем $\approx 0,7$ [Башмаков, 2005]. Кроме того, современные системы классификации узкоспециализированных полнотекстовых документов, используемые в отраслях, применяют статические системы классификации, со строго заданными иерархическими отношениями между тематическими классами. В случае расширения предметных областей и развития знаний, точность и полнота выдачи таких систем резко падает и требует ручного изменения классификационной схемы [Бабенко, 2004].

Выводы

Результатом данного исследования является математическая модель отношений между областью интеллектуальной деятельности менеджера и достоверными глубинными знаниями, представленными в документах. Реализация модели дает возможность осуществлять динамическое разбиение поступающих в КИС документов и содержащихся в них терминологических понятий. Использование данной модели позволило разработать информационное и программно-алгоритмическое обеспечение семантико-ориентированного лингвистического процессора, основанного на работе логических сетей, извлекающего новые терминологические понятия из поступающих в систему документов и относящего их к классам, соответствующим областям деятельности различных менеджеров. Извлекаемые понятия наполняют новой информацией базы знаний корпоративной информационной системы, делая ее динамичной и одновременно увеличивая семантическую силу модели представления знаний КИС. Приведенные расчеты эффективности программной реализации модели показывают достаточно высокую полноту и релевантность системы разбиения документационных потоков по соответствующим областям деятельности менеджеров КИС.

Литература:

- [Carroll, 2005] Carroll, J., Evans, R., and Klein, E. Supporting Text Mining for e-Science: the challenges for Grid-enabled Natural Language Processing. UK e-Science All Hands Meeting, Nottingham, UK, 2005.
- [Cormack, 1998] Cormack G.V., Palmer C.R., Clarke C.L.A., Efficient construction of large test collections // Proc. of the SIGIR'98 - pp. 282-289.
- [Khairova, 2009] Nina Khairova, Natalia Sharonova. Use of Predicate Categories for Modelling of Operation of the Semantic Analyzer of the Linguistic Processor./Proceedings of IEEE EAST-West Design & Test Symposium (EWDTS'09). // Moscow, Russia, September 18-21, 2009/ — P. 204- 207
- [Mancini, 2001] Mancini J. Enterprise Content Management: Critical Technologies for Business Applications // AIIM, 2001.
- [Mizzaro, 1996] Mizzaro S. How many relevances in information retrieval?, in C.W. Johnson and M. Dunlop (eds), Proceeding of the Workshop 'Information Retrieval and Human Computer Interaction', GIST Technical Report GR96-2, Glasgow University, The British Computer Society, Glasgow, UK, pp. 57-60.
- [Бабенко, 2004] Система классификации текстов информационных сообщений на русском языке "АКТИС"/ Бабенко М. и др. // Труды международной конференции "Программные системы: теория и приложения", ИПС РАН, г. Переславль-Залесский, май 2004 — М.: Физматлит, 2004, Том 2., с. 7-21.
- [Башмаков, 2005] Башмаков А. И., Башмаков И. А. Интеллектуальные информационные технологии. — М.: Изд-во МГТУ им. Н.Э.Баумана, 2005. — 304 с.
- [Бондаренко, 2007] Бондаренко М.Ф., Шабанов-Кушнарченко Ю.П. Теория интеллекта. Харьков. Изд-во СМИТ. 2007. 576 с
- [Гаврилова, 2000] Гаврилова Т. А. Базы знаний интеллектуальных систем. – СПб : Питер, 2000.
- [Кураленок, 2002] Кураленок И., Некрестьянов И., Оценка систем текстового поиска. // Программирование. 2002, 28(4):226-242.
- [Хайрова, 2003] Хайрова Н.Ф., Шаронова Н.В. Автоматизированные информационные библиотечные системы: задачи обработки информации: Монография. — Х.: Нар укр. акад., 2003. — 120 с.
- [Шабанов-Кушнарченко, 1986] Шабанов-Кушнарченко Ю.П. Теория интеллекта: Технические средства. — Х.: Вища школа, 1986.— 134с.
- [Шабанов-Кушнарченко, 1993] Шабанов-Кушнарченко Ю.П., Шаронова Н.В. Компараторная идентификация лингвистических объектов: Монография.— К.: ИСДО, 1993.— 116 с.

Сведения об авторах

Хайрова Нина – доцент кафедры интеллектуальных компьютерных систем Национального технического университета «Харьковский политехнический университет», ул. Фрунзе, 21, Харьков, 61002, Украина e-mail: nina_khajrova@yahoo.com

Научные интересы: искусственный интеллект, обработка знаний, автоматическая обработка текстов

Шаронова Наталья – профессор, заведующий кафедрой интеллектуальных компьютерных систем Национального технического университета «Харьковский политехнический университет», ул. Фрунзе, 21, Харьков, 61002, Украина e-mail: nvsharonova@mail.ru

Научные интересы: искусственный интеллект, математическое моделирование, автоматизированные библиотечные системы

МОДЕЛЬ ВИРТУАЛЬНОГО МИРА МУЛЬТИМЕДИА ТРЕНАЖЕРА ДЛЯ МЕДИЦИНСКОГО ОБРАЗОВАНИЯ

В.В. Грибова, М.В. Петряева, Л.А. Федорищев, М.Ю. Черняховская

Абстракт: В работе представлена модель виртуального мира медицинского мультимедиа тренажера. Описаны компоненты модели: объекты, действия, которые смогут совершать обучаемые с объектами виртуально мира. Приведены фрагменты сформированных сценариев обучающих заданий.

Ключевые слова: виртуальный мир, медицинские компьютерные обучающие тренажеры, объекты, сценарии.

АСМ-классификация: D.2.2 [Технология программирования]: Методы и средства проектирования D.2.10 [Технология программирования]: Проектирование---методология; H.1.1 [Модели и принципы]: Теория систем и информации---общая теория систем.

Введение

В современной образовательной практике компьютерные тренажеры стали очень востребованными для профессиональной подготовки специалистов разного уровня квалификации [Гаммер М.Д., 2007]. До недавнего времени в процессе подготовки медицинских кадров превалировал метод обучения «на больном», т. е. на реальном объекте. Однако в настоящее время в странах Евросоюза запрещено обучение студентов на реальных больных, в России требуется обязательное согласие пациента на такие формы обучения [Мещерякова М.А., 2006]. Решение этой проблемы можно добиться путем использования в учебном процессе обучающих тренажеров, которые позволят существенно улучшить качество обучения, дадут возможность обучаемым "проиграть" различные практические ситуации до начала их работы в реальных условиях, обеспечат возможность удаленного обучения [Шубина Л.Б. и др., 2010]. Разработка медицинских компьютерных тренажеров с использованием мультимедиа технологий на базе технологий виртуального окружения, создает возможность реализовать практически любые исследования на виртуальном пациенте и воспроизвести методики отработки многих медицинских манипуляций. В итоге, как студент, так и начинающий врач, получают не только теоретическую, но и высококачественную практическую подготовку.

В работе [Грибова В.В. и др., 2009] на основе анализа тренажеров, описанных в литературе, была предложена концепция их разработки. В соответствии с предложенной концепцией выделены следующие компоненты модели тренажера:

- **модель виртуального мира**, которая состоит из описания объектов, команд, действий и сценариев проводимых исследований;
- **модель диагностических знаний**, состоящая из наблюдений, диагностируемых состояний, причинно-следственных связей между наблюдениями и диагностируемыми состояниями;
- **модель упражнений**, которая включает различные диагностические ситуации (примеры), приемы медицинских манипуляций и т.п.;
- **модель оценки действий обучаемого**, включающая оценку результатов проведения диагностических исследований и объяснение ошибок, совершенных обучаемым при их проведении.

Каждому компоненту модели тренажера соответствует онтология, которая описывает информацию о соответствующем компоненте модели. Формирование конкретного компонента модели тренажера сводится к выделению подмножества из соответствующей онтологии и уточнении значений ее характеристик.

Целью настоящей работы является описание модели виртуального мира медицинских компьютерных обучающих тренажеров, которые планируется использовать для отработки профессиональных знаний, умений и навыков у студентов медицинских вузов в процессе обучения их врачебному искусству (делу).

Компоненты виртуального мира медицинского компьютерного обучающего тренажера

Виртуальный мир медицинского компьютерного обучающего тренажера строится в соответствии с критериями идеализации и абстракции; идеализация состоит в том, чтобы отобразить в виртуальном мире существенные для него признаки объектов, а абстракция — в том, чтобы игнорировать другие, нерелевантные для данного виртуального мира особенности (объектов) [Филатова Н.Н., Вавилова Н.И., 2000]. Виртуальный мир является трехмерным и действие происходит в режиме реального времени.

Компонентами виртуального мира являются: множество объектов, команд, действий и сценариев обучающих заданий.

Характеристика объектов виртуального мира

Объекты имеют разную структуру, роль, поведение и отображение. Каждый объект обладает набором *атрибутов*, определяющих присущие ему свойства, положение в пространстве и т.д. Исходя из этого, выделены следующие основные **классы** объектов (см. рис. 1).

- **Простой (неизменяемый) объект.**
- **Изменяемый объект.**
- **Составной объект.**
- **Таблица (массив, регулярная структура).**

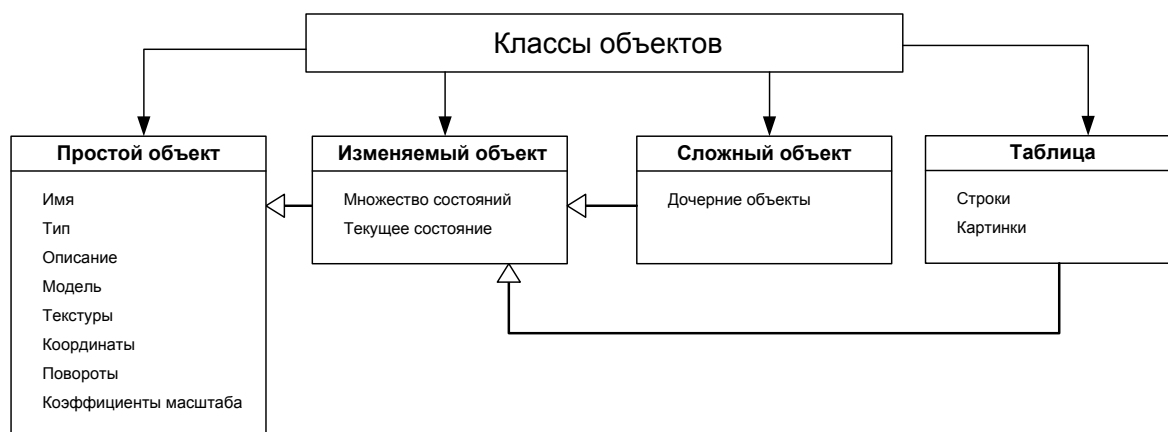


Рис. 1. Схема наследования классов объектов.

Простой (неизменяемый) объект – это объект, обладающий набором атрибутов, которые не изменяются в процессе воспроизведения сцены, например стол, кушетка. Простой объект можно назвать "базовым" в силу того, что от него наследуются все остальные типы объектов. Кроме того, простой объект обладает набором обязательных атрибутов *логического уровня* и *уровня представления*:

- имя – уникальный идентификатор объекта, термин предметной области.
- тип – атрибут, определяющий принадлежность данного объекта к определенному классу объектов.

- описание – текстовое описание объекта, которое делает эксперт предметной области для пояснения его предназначения.
- модель - содержит ссылку на файл к полному описанию 3d-модели.
- текстуры (текстура) – это [растровое изображение](#), накладываемое на одну из поверхностей модели для придания ей цвета, окраски или иллюзии рельефа. Координаты - три вещественных числа (координаты x, y, z), которые определяют положение объекта на сцене в мировой системе координат.
- углы вращения - три вещественных числа, углы вращения вокруг осей x, y, z (от нуля до 360 градусов). Это независимые числа и могут быть разными в разных состояниях объекта.
- коэффициенты масштабирования - три вещественных числа, множители по осям x, y, z , которые необходимы для изменения размеров объекта.

Углы вращения могут быть заданы, если необходимо изменить ориентацию модели. На основе образца модели может быть добавлено множество экземпляров этой модели, имеющих разную пространственную ориентацию. Аналогично для коэффициентов масштабирования: модели могут быть одинаковыми с точностью до размеров (например, одни стулья могут быть больше, другие меньше).

При описании объектов виртуального мира предметной области «Медицина» можно выделить следующие простые объекты: "кушетка", "стол врача", "стул" - они являются простыми "декорациями", которые не изменяются в процессе работы приложения. Для них достаточно определить, как они выглядят и где находятся в сцене виртуального мира. Далее, в качестве примеров, будут приведены фрагменты формального описания различных классов объектов, сформированные экспертами предметной области для офтальмологического медицинского обучающего тренажера.

Пример 1. Формальное описание объекта "Стол медицинский".

- имя: "стол медицинский",
- тип: простой,
- описание: "предмет мебели, состоящий из горизонтальной поверхности (столешницы) и основания, предназначенный для размещения медицинского инструментария",
- модель: "стол",
- координаты: [100, 100, 0],
- текстуры, углы вращения, коэффициенты: можно в данном случае не задавать.

Изменяемый объект – это объект, который является расширением неизменяемого объекта, т.е. обладает всеми обязательными атрибутами, которые есть у простого объекта, а также включает множество *состояний*, которые изменяют его атрибуты в процессе воспроизведения сцены:

- множество состояний - множество изменяемых атрибутов объекта, которые есть у данного класса объектов.
- текущее состояние - содержит ссылку (название) на состояние из представленного множества всех возможных состояний.

В описании объектов виртуального мира предметной области "Офтальмология. Проверка зрения" можно выделить следующие изменяемые объекты: "окклюдер", «скиаскоп», "аппарат Рота" и др. Объект «окклюдер» имеет одну и ту же модель и текстуры, следовательно, эти атрибуты у него – постоянные. С другой стороны, по сценарию исследования пациента, окклюдер изменяет свои состояния: сначала лежит на столе, затем находится в руке пациента, затем закрывает левый глаз пациента, затем правый глаз пациента и тд. Кроме того, объект «окклюдер» в течение работы может иметь разные координаты и повороты, которые, соответственно, должны отображаться в его состояниях.

Пример 2. Формальное описание объекта "Окклюдер".

- имя: "окклюдер",
- тип: изменяемый,
- описание: "ручной щиток для прикрытия одного глаза при проверке остроты зрения другого глаза",
- модель: "окклюдер",
- множество состояний: [

«лежит на столе»: {координаты: [1,2,3], повороты: [90,0,0]},

«в руке пациента»: {координаты: [4,5,6], повороты: [45,0,45]},

«в руке пациента, закрывает левый глаз»: {координаты: [7,8,9], повороты: [0, 90, 0]},

«в руке пациента, закрывает правый глаз»: {координаты: [7,8,9], повороты: [0, 90, 0]},

- текущее состояние: «в руке пациента, закрывает левый глаз»...

Составной объект – это объект, который наследуется от изменяемого объекта, обладает всеми его атрибутами и, кроме этого, может содержать *дочерние* объекты, образуя иерархию.

Для образования иерархической структуры у родительского объекта должен быть добавлен дополнительный атрибут-указатель на дочерние объекты. Этот атрибут может быть изменяемым (находиться на уровне состояний). У объекта может быть указано несколько дочерних объектов. Дочерние объекты могут быть объектами любых типов. У всех дочерних объектов (или подобъектов) обязательно есть состояния, соответствующие состояниям родительского объекта (могут быть и свои уникальные состояния, но те, которые есть у родительского объекта, должны быть в обязательном порядке). Такая иерархическая связь обеспечивает управление всем сложным объектом, как единым целым. Таким образом, если, например, у родительского объекта изменить состояние, то, «спускаясь вниз» по всем дочерним объектам и находя состояние с таким же названием, которое было выбрано для родительского объекта, будут меняться состояния этого дочернего объекта. В рассматриваемом примере из предметной области "Офтальмология. Проверка зрения" составным объектом является *пациент*. Пациент, с одной стороны, является единым целым объектом, у которого есть координаты в пространстве, модель, набор состояний; с другой стороны – он является составным из нескольких других объектов - частей: руки, ноги, голова и т.д., каждый из которых тоже может быть объектом любого типа (включая составной), и каждый из них имеет те же состояния, что и объект-целое. Пациент в данном примере имеет следующие постоянные атрибуты: модель (не включая модели рук и головы), текстуры, координаты (если он только сидит на стуле, в противном случае, они могут быть изменяемыми).

Пример 3. Формальное описание объекта "Пациент".

- имя: "пациент",
- тип: составной,
- описание: "человек, подвергающийся медицинскому обследованию, наблюдению и/или лечению",
- модель: "человек",
- координаты: [5,10,0],
- множество состояний: [
 - "сидит, руки на коленях, без окклюдера": {...},
 - "сидит, держит окклюдер левой рукой, правая рука на коленях": {...},
 - "сидит, держит окклюдер правой рукой, левая рука на коленях": {...}],
- текущее состояние: «сидит, руки на коленях, без окклюдера»,
- дочерние объекты: ["рука левая", "рука правая", "голова", ...].

Такие части пациента как руки, тоже являются сложными объектами, они тоже имеют дочерний объект "окклюдер" в своих состояниях и описываются следующим образом.

Пример 4. Формальное описание дочернего объекта "Рука левая".

- имя: "рука левая",
- тип: составной,
- описание: "управляемая часть пациента",
- модель: "рука",
- множество состояний: [
 - "сидит, руки на коленях, без окклюдера": {координаты: [1,2,3], повороты: [0, 0, 0]},
 - "сидит, держит окклюдер в левой руке, правая рука на коленях": {координаты: [4,5,6], повороты: [0, 0, 0], дочерний объект: "окклюдер"},
 - "сидит, держит окклюдер в правой руке, левая рука на коленях": {координаты: [7,8,9], повороты: [0, 90, 0]}],
- текущее состояние: «сидит, руки на коленях, без окклюдера»...

Таблица (массив, регулярная структура) – это объект, имеющий линейно (в нескольких измерениях) упорядоченную структуру. Таблица наследуется от изменяемого объекта и расширяет изменяемый объект. Таблица имеет следующие специфические атрибуты:

- строки – ряды элементов,
- картинки - набор шаблонов картинок, соответствующих элементам таблицы.

В рассматриваемом примере из предметной области "Офтальмология. Проверка зрения" присутствует объект *таблица* - «Таблица для определения остроты зрения». Таблица во время исследования находится в одном и том же месте, имеет одну и ту же модель, следовательно - это неизменяемые атрибуты. Текстура для нее генерируется на основе введенных данных (которые также являются неизменяемыми атрибутами на уровне объекта) и является параметром изменяемым, так как может быть освещенной или неосвещенной.

Пример 5. Формальное описание объекта "Таблица для определения остроты зрения"

- имя: "таблица для определения остроты зрения Сивцева",
- тип: таблица,
- описание: "таблица, состоящая из 12 рядов оптотипов «буквы»",
- модель: "плакат",
- координаты: [5, 5, 0],
- строки: [

1-ая строка {элементы: ["буква Ш", "буква Б"], высота: 70 мм},

2-ая строка {элементы: ["буква М", "буква Н", "буква К"], высота: 35 мм},

3-ья строка {...}

...

- картинки: {"буква Ш", "буква Б", "буква М", "буква Н", "буква К", ... }, множество состояний: {"освещена», «неосвещена»},
- текущее состояние: «неосвещена»...

Характеристика действий виртуального мира тренажера

Основное назначение действий – это получение, отображение и передача результатов, связанных с изменением виртуальной среды системы. Действия могут быть следующих основных типов по виду выполнения:

- **Действие командное** – это предписание объектам виртуальной среды произвести какое-либо действие без взаимодействия пользователя с ними, например «взять окклюдер в левую руку», «закрыть левый глаз окклюдером» и т.п.

- **Действие интерактивное** – это такое действие, которое пользователь выполняет сам, через взаимодействие с объектами виртуальной среды, например «включить аппарат Рота», «надеть пробную оправу на пациента» и т.п.

Действия обладают следующим набором атрибутов:

- имя - уникальный идентификатор;
- тип - командное или интерактивное действие;
- описание - эксперт описывает свое представление о действии;
- входные параметры (необязательный атрибут) – описываются дополнительные данные, влияющие на выполнение действия;
- состояния объектов (необязательный атрибут) - описывает множество состояний различных объектов, которые участвуют в действии; если при выполнении действия состояние объекта изменяется, то необходимо описать, какой объект и какое его состояние должно быть выбрано;
- результат (необязательный атрибут) - множество возможных ответов (результатов, "оценок результатов"), которые могут быть получены после выполненного действия и имеющие значение для сценария.

Каждый результат, возвращаемый действием, состоит из двух частей:

- *отображения результата в виртуальной среде* – полученный результат отображается пользователю в виртуальной среде после завершения действия. Например, если пользователь спрашивает пациента, где находится разрыв кольца, тогда отображением результата будет его ответ "сверху", "снизу", "справа" или "слева";
- *оценки результата для сценария, которая* передается для анализа в сценарий обучающих заданий (см. далее) и не отображается пользователю. Например, для интерактивного действия "назвать буквы на 10-ой строке таблицы" оценкой результата для сценария могут быть "видит" или "не видит" пациент буквы на строке таблицы.

Пример 6. Формальное описание действия "Закрыть левый глаз окклюдером"

- имя: "закрыть левый глаз окклюдером",
- тип: командное действие,
- описание: "при указании данной команды пациент должен закрыть левый глаз окклюдером",
- состояния объектов: {объект: "Пациент", состояние: "левый глаз закрыт"} (при условии, что в модели уже описан объект "Пациент" со всеми состояниями).

Пример 7. Формальное описание действия "Назвать буквы на строке"

- имя: "назвать буквы на строке",
- тип: интерактивное действие,
- описание: «пользователь щелкает мышкой по строке на таблице, чтобы пациент назвал буквы на этой строке»,
- входные параметры: ["номер строки" (число)],
- результат: ["видит", "не видит", "ошибка"].

Формирование сценариев обучающих заданий

При работе с тренажером пользователь должен правильно выполнить обучающее задание и в требуемой последовательности. Для этого должен быть написан сценарий, в соответствии с которым будут проверяться действия пользователя. Сценарий представляет собой граф, в узлах которого находятся

действия, а дуги являются переходами от одних действий к другим. Каждый переход означает возможную альтернативу выбора следующего действия. Переходы могут образовывать циклы (как прямые, так и цепочные).

Выделены следующие виды переходов:

- **Фиксированный переход.**
- **Переход, определяемый результатом действия.**
- **Переход, определяемый условием.**

Фиксированный переход может использоваться для простого упорядочивания последовательности действий (см. рис.2).

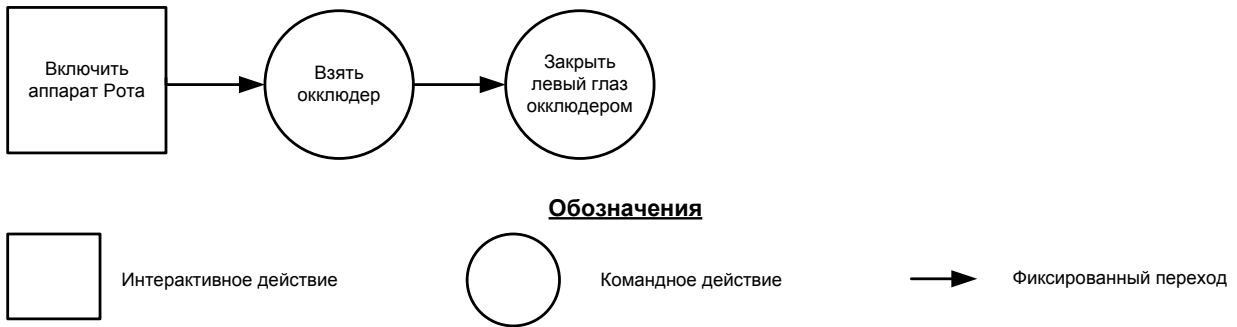


Рис. 2. Фрагмент сценария с фиксированными переходами.

Переход, определяемый результатом действия зависит от полученных результатов выполненных действий. На основе полученного результата для сценария из текущего действия делается решение о переходе к следующему действию. Допустим, по сценарию обучающего задания необходимо проверить, как пациент видит строки таблицы для определения остроты зрения. Для этого пользователь просит назвать пациента буквы на различных строках таблицы и в зависимости от его ответов принимает решение о следующих проверках (см. рис. 3).

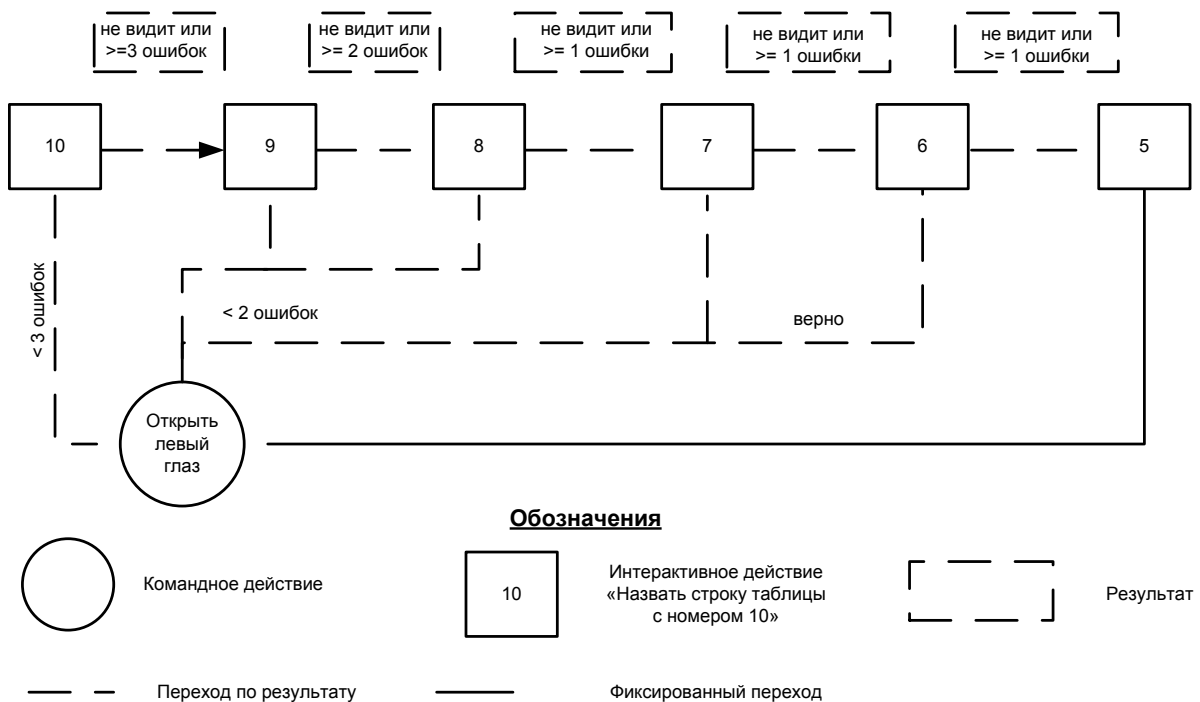


Рис. 3. Пример сценария с переходами по результату ответов пациента.

Переход, определяемый условием. Условие – это структура из трех компонентов:

- атрибут объекта, который проверяется в условии. Например, "оптотип расстояние". Здесь будет проверяться атрибут "расстояние" у объекта "оптотип".
- операция сравнения – это любая логическая операция по сравнению двух величин. Например, такие операции как "больше" (" $>$ "), меньше (" $<$ ") и т.д.
- значение – проверяемая величина, с которой сравнивается атрибут. Данное значение может быть как терминальной величиной, так и вычисляемой.

На рис. 4 представлен пример сценария, который включает переходы по условию. Этот пример является наиболее близким к реальному сценарию. В нем присутствуют все описанные виды переходов между командными и интерактивными действиями.

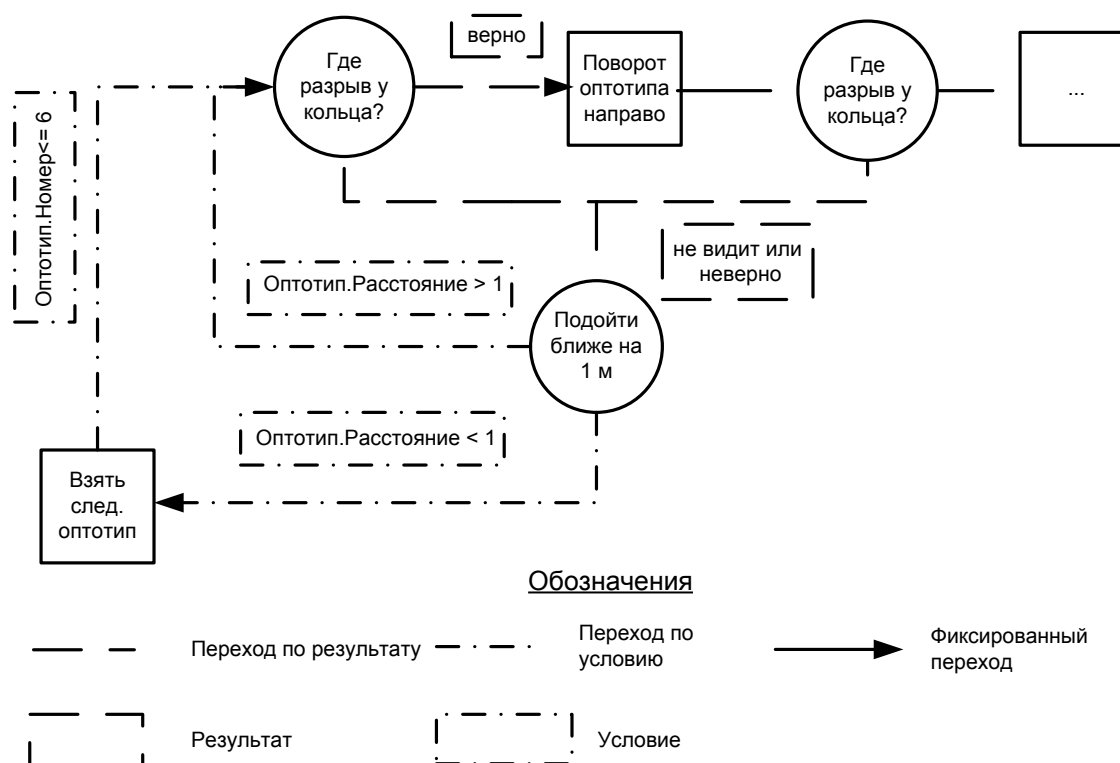


Рис. 4. Фрагмент сценария с различными видами переходов.

Таким образом, описанные переходы дают возможность эксперту моделировать сценарии обучающих заданий любой сложности.

Заключение

В работе представлена модель виртуального мира компьютерного обучающего тренажера. С использованием описанной модели разработана экспериментальная версия компьютерного обучающего тренажера для раздела медицины «Офтальмология», которая подтвердила корректность описания модели. В настоящее время продолжают работы по описанию сценариев для различных методов исследования в офтальмологии на основе представленной модели и их реализация. Компьютерный обучающий тренажер планируется использовать для обучения студентов медицинских вузов.

Благодарности

Работа выполнена при финансовой поддержке РФФИ, грант «Управление сообществами интеллектуальных систем» и ДВО РАН, грант «Разработка теоретических основ инструментальной поддержки проектирования, генерации и отладки диагностических компьютерных тренажеров».

Библиография

1. Гаммер М.Д. Применение компьютерных имитационных тренажеров и систем виртуальной реальности в учебном процессе // <http://cde.tsogu.ru/publ1/>.
2. Мещерякова М.А. Учебный процесс вуза в системе управления качеством профессиональной подготовки врачей. Монография. М.: КДУ, 2006. - 140 с.
3. Шубина Л.Б. Мещерякова М.А., Камынина Н.Н., Уткина Г.Ю. Развитие медицинского образования в условиях инновационной экономики <http://vestnik.mednet.ru/content/view/182/30/>.
4. Филатова Н.Н., Вавилова Н.И. Проектирование мультимедиа тренажеров на основе сценарных моделей представления знаний // Educational Technology S Society 3(4), 2000.- P.193-202.
5. Грибова В.В., Осипенков Г.Н., Сова С.А. Концепция разработки диагностических компьютерных тренажеров на основе знаний // International Book Series "Human Aspectsof Artifical Intelligence". N12. Papers are selected from Proc. of the Intern. Conf. of the Join International Events of Informatics "ITA 2009" (e.TECH-2009), Varna, Bulgaria, 2009. P. 27-33. – Suppl. to the International Journal "Information Technologies & Knowledge". 2009. Vol.32. Sofia, Bulgaria (Inst. of Information Theories and Applications FOI ITHEA)
6. V. Gribova. The concept of an intelligent tool for development of diagnostic computer simulators // Proceedings of First Russia and Pacific Conference on Computer Technology and Applications (RPC 2010). [Electronic res.]. Vladivostok (Russia), 6-9 September 2010. -Vladivostok: IACP FEB RAS, 2010. - ISBN: 978-0-9803267-3-4 (CD). Pp.63-65.

Информация об авторах



Грибова Валерия Викторовна - д.т.н., зав. лабораторией интеллектуальных систем Института автоматки и процессов управления ДВО РАН, 690041, Россия, г. Владивосток, ул. Радио, 5 Тел: +7 4232 314 001 E-Mail: gribova@iacp.dvo.ru
Область научных исследований: искусственный интеллект, интеллектуальные системы, пользовательские интерфейсы



Петряева Маргарита Вячеславовна – к.м.н., н.с. лаборатории интеллектуальных систем Института автоматки и процессов управления ДВО РАН, 690041, Россия, г. Владивосток, ул. Радио, 5 Тел: +7 4232 314 001 E-Mail: margaret@iacp.dvo.ru
Область научных исследований: медицинские экспертные и информационные системы, формализация медицинских знаний.



Федорищев Леонид Александрович – м.н.с. лаборатории интеллектуальных систем Института автоматки и процессов управления ДВО РАН, 690041, Россия, г. Владивосток, ул. Радио, 5 Тел: +7 4232 314 001 E-Mail: fleo1987@mail.ru
Область научных исследований: искусственный интеллект, виртуальная реальность.



Черняховская Мери Юзефовна – д.м.н., гл.н.с. лаборатории интеллектуальных систем Института автоматки и процессов управления ДВО РАН, 690041, Россия, г. Владивосток, ул. Радио, 5 Тел: +7 4232 314 001 E-Mail: chernyah@iacp.dvo.ru
Область научных исследований – медицинская информатика

МОДЕЛЬ ДИНАМИКИ МОНЕТНОГО ОБРАЩЕНИЯ В СРЕДНЕВЕКОВОЙ ПЛИСКЕ

Йордан Табов, Галина Панайотова

Абстракт: В настоящей работе представлена модель динамики монетного обращения в старой болгарской столице Плиске, основанная на найденных старых монетах. Если найденные там монеты датированы правильно, анализ изменения объема монетного обращения приводит к выводу, что Плиска была „городом” (в настоящем смысле этого слова) только в период примерно с 970 до около 1100 года.

Ключевые слова: модель, монетное обращение, монетные находки, Плиска

ACM Classification Keywords: I.6 SIMULATION AND MODELING, I.6.3 Applications

Введение: Можно ли по количествам найденных старых монет, отчеканенных в разные периоды прошлого, судить об изменениях объема монетного обращения?

Положительный ответ (с некоторыми оговорками) на этот вопрос предложен в 1954 г. А. Кажданом в его исследовании о развитии византийских городов [Каждан, 1954], где он между прочим пишет: „... обилие монет, добытых при раскопках того или иного города, может служить серьезным аргументом, доказывающим здесь наличие товарного производства” [Каждан, 1954]. В этом исследовании Каждан использует в качестве объективного археологического материала старые византийские монеты. Он пользуется **двумя видами источников информации**: 1) **каталогами** больших музейных и частных коллекций, и 2) **публикациями о находках** (раскопок или сокровищ). Каждан подчеркивает, что **анализ информации о достаточно большом количестве монет дает возможность смоделировать интенсивность чеканки монет** [Каждан, 1954] (**а исходя из него – изменения объема монетного обращения) в прошлом.**

Несколько лет после публикации статьи Каждана, снова советский историк – И. В. Соколова [Соколова, 1958] – предлагает другое объяснение почему от разных периодов прошлого до нас дошли разные количества монет. Она считает, что большинство найденных в наше время старых монет были скрыты их бывшими собственниками из-за военной опасности (т.е. из-за страха от захватчиков) [Соколова, 1958].

А. Гандила [Gândilă, 2009] указывает на Д. М. Меткалфа [Metcalf, 1958 и Metcalf, 1960] как на одного из пионеров в применении статистических методов для исследования монетного обращения в Византийской империи и добавляет еще имена В. Е. Меткалфа и С. Морриссона [Morrisson, 1980 и Morrisson, 1981]. По его оценке, после 1980 г. наблюдается настоящая „эксплозия” в использовании более или менее сложных статистических подходов.

Но специально для обоснования утверждения, что **количество найденных старых монет из определенного периода прошлого может служить индикатором для объема монетного обращения**, особое место занимает исследование Дугласа Ньютона [Newton, 2006]. Сравнивая дошедшую до нас информацию о количествах отчеканенных в прошлом монет с данными археологических раскопок он установил корреляцию с большим коэффициентом Пирсона (около 0.9) между количествами „выпущенных” и „найденных” монет [Newton, 2006]. Она дает возможность „измерять” изменение хозяйственной активности в данном городе, регионе или государстве путем применения количественного анализа найденных на соответствующей территории старых монет.

О монетах, найденных в Плиске

Настоящее исследование основывается на данных, опубликованных в статье И. Йорданова [Йорданов, 2000], дающей полную наличную информацию о старых монетах, найденных в Плиске и ее окрестности. Общее число охваченных в исследовании монет превосходит 550 экземпляров.

период	медные	золотые	серебряные	билоновые	всего
117-346	6	-	-	-	6
IV-Vв.	14	-	-	-	14
491-518	2	-	-	-	2
518-527	1	-	-	-	1
527-565	4	-	-	-	4
565-578	2	-	-	-	2
V-VI	5	-	-	-	5
802-811	2	-	-	-	2
867-886	1	2	-	-	3
886-912	8	-	-	-	8
913-919	1	-	-	-	1
931-944	2	-	-	-	2
945-950	1	-	-	-	1
945-959	-	2	-	-	2
963-969	4	-	-	-	4
969-976	37	-	1	-	38
976-1025	-	2	-	-	2
976-1035	365	-	2	-	367
1028-1034	-	1	-	-	1
1034-1041	-	1	-	-	1
1035-1041	74	-	-	-	74
1042-1050	9	-	-	-	9
1042-1055	-	-	1	-	1
1057-1059	-	1	-	-	1
1050-1060	9	-	-	-	9
1059-1067	8	1	-	-	9
1075-1080	1	-	-	-	1
1092-1118	-	-	-	1	1
всего	556	10	2	1	569

Рис 1. Таблица числа монет, найденных в Плиске и ее окрестности, по видам (медные, золотые, серебряные и билоновые) и по периодам чеканки [Йорданов, 2000].

Интересующая нас информация о числе монет, найденных в Плиска и ее окрестности, по видам (медные, золотые, серебряные и билиновые) и по времени чеканки, в указанной статье Йорданова систематизирована в таблице, которая приведена здесь на **рис. 1**.

Первый вывод, который вытекает из чисел в таблице, состоит в том, что почти все учтенные в ней монеты (около 98 %) – медные. Т.е. это монеты, которые участвуют в повседневном обращении, на них покупают хлеб, молоко, вино, овощи; среди них очень мало золотых, серебряных и билиновых монет, которые имеет смысл собирать и хранить в тайных местах.

Распределение найденных в Плиске монет по векам

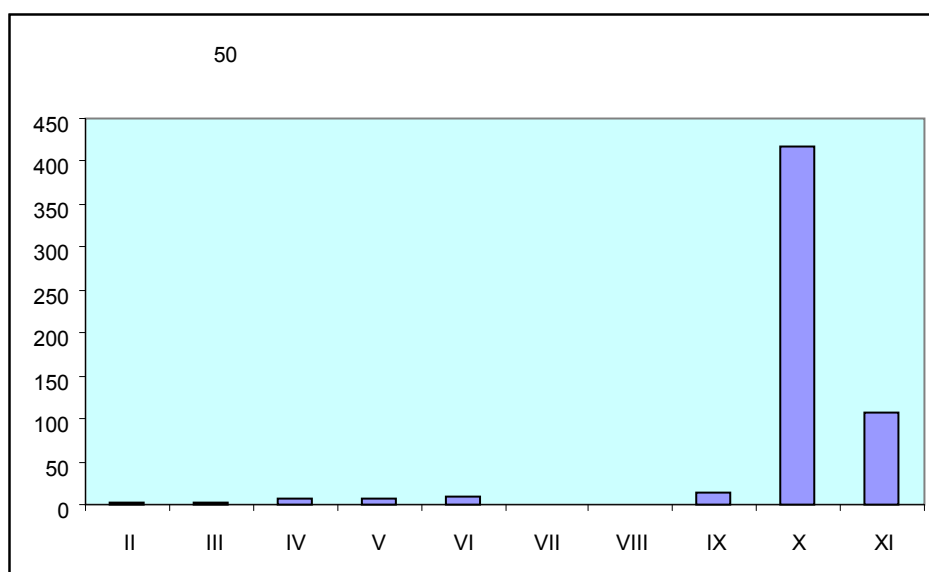


Рис. 2. Диаграмма распределения найденных в Плиске монет по векам.

Благодаря работе ряда поколений ученых-нумизматов, для почти всех найденных старых монет определено во время какого царя (или правителя) их чеканили. Это дает приближенную – обычно с точностью до нескольких лет или немного больше – датировку их поступления в обращение. Таким образом, старые монеты данной совокупности можно сгруппировать «по владельцам», в каждой полученной группе монеты нужно подсчитать, данные можно упорядочить в таблице, подобной таблице на рис. 1, и проанализировать их.

Недостаток этого подхода состоит в том, что длины «периодов царствования» владельцев иногда очень сильно (в 10-20 раз и более) отличаются друг от друга.

Чтобы преодолеть этот недостаток можно сгруппировать монеты по длинным интервалам, например по 100 лет. Распределение совокупности старых монет **по векам** можно сделать без особого труда. Для найденных в Плиске монет это приводит к результату, представленному в наглядном виде на **рис. 2**.

Устранения указанного недостатка можно добиться путем распределения монет по одинаковым коротким интервалам, например по 1 году. Такой подход дает хорошую точность, но при работе с большим числом монет занимает неоправданно много времени.

Для целей исторического исследования изменения хозяйственной и финансовой активности обычно достаточно использовать интервалы не по одному году, а несколько большей длины – например по 10 или даже 20 лет; такие мы предпочтем ниже. До этого однако мы воспользуемся картиной, представленной на **рис. 2**: из нее видно, что от веков с II по VI в Плиске найдено очень мало монет, а с VII и VIII века таких среди находок и вовсе нет. Поэтому в дальнейшем мы сосредоточимся на монетах с IX, X и XI веков.

Хронологическое распределение найденных в Плиске монет

Понятие хронологического распределения информации в исторических текстах (основанное на введенной А. Фоменко “функции объема”) [Фоменко, 1981а; Фоменко, 1981b; Fomenko and Rachev, 1990], методика его построения и возможные обобщения предложены в статье Й. Табова [Tabov, 2003]. Вариации этого понятие и методики его построения применялись для конструирования и использования специфических “хронологических распределений” монет [Tabov, Vasilev and Velchev, 2003], старых рукописей [Tabov, J., A. Velchev, M. Dobрева, K. Sotirova, 2004] и музейных экспонатов [Hristova and Dobрева, 2004].

Здесь мы применяем фактически ту же методику для построения функции самого хронологического распределения и соответственно его графика. Для этого необходимо располагать данными о **приблизительной** датировке каждой монеты. Учитывая особенности целей, для которых мы используем хронологическое распределение, достаточно построить его приблизительно; например, построить приблизительно **представляющий его график, так чтобы отчетливо выделялись его максимумы и минимумы, интервалы возрастания и убывания**. Изменения графика в узких границах могут иметь (и чаще всего имеют) случайный характер и не должны оказывать влияние на выводы проистекающие из свойств графика.

Хронологическое распределение найденных в данной стране, регионе или городе монет может служить **моделью, представляющей динамику** (изменение – возрастание и убывание) **объема монетного обращения в разные периоды времени**. Установленная Д. Ньютоном высокая степень корреляции количеств отчеканенных в прошлом монет данного вида и найденных в наши дни монет того же вида показывает, что чем большая случайная выборка найденных монет использована для построения их хронологического распределения, тем точнее модель, и с тем большей степенью надежности можем использовать его для обоснования конкретных выводов об исторической картине прошлого.

Диаграмма на **рис. 2** по существу отражает хронологическое распределение найденных в Плиске старых монет по данным И. Йорданова [Йорданов, 2000] с базовым интервалом длиной в 1 век (100 г.). Но она весьма „грубая”. Если выберем базовый интервал из 10 лет (вместо 1 века) и обработаем данные Excel - ом, получаем более точный график, часть которого – для периода с 800 г. до 1100 г. – представлена на **рис 3**.

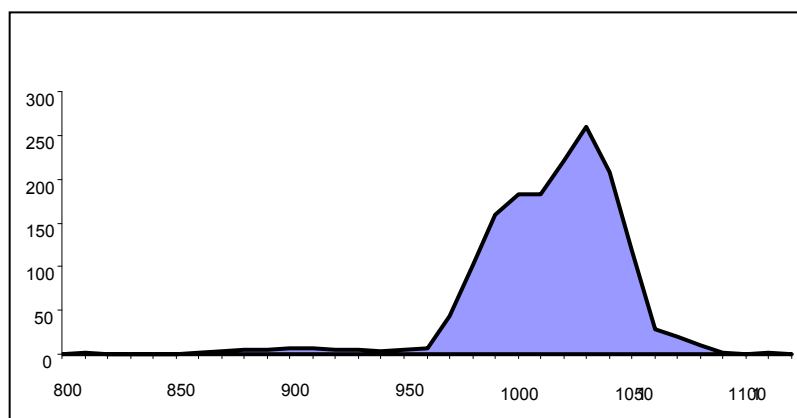


Рис. 3. Хронологическое распределение найденных в Плиске монет, выпущенных в период 800-1120 г.

Улучшенная модель динамики монетного обращения в Плиске

График на **рис. 3** показывает **приблизительное распределение** монет по годам выпуска; но обычно каждая монета участвует в монетном обращении несколько лет – пока не испортится, не потеряется или не будет выведена из оборота. Мы принимаем, что средняя продолжительность участия монет из рассматриваемой здесь совокупности в обращении составляет 20 лет. Тогда если через $f(x)$ обозначить функцию, определенную графиком на **рис. 3**, можем получить лучшую модель динамики монетного

обращения. Это можно сделать так: вместо функции $f(x)$ хронологического распределения на **рис. 3** возьмем сумму $g(x) = f(x) + f(x - 20)$; эта модель в нашем случае представлена графически на **рис. 4**.

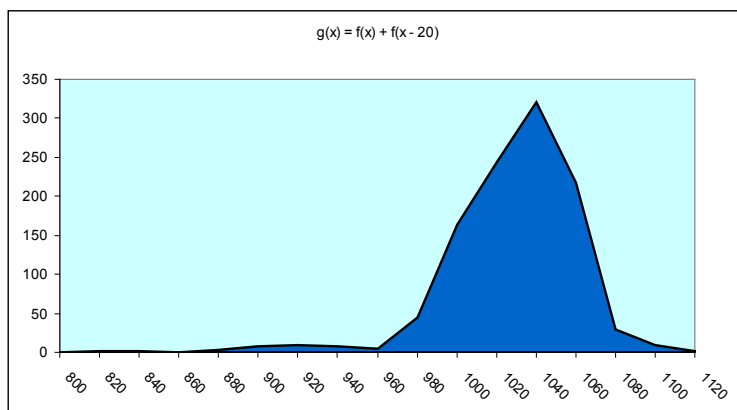


Рис. 4. Модель изменения объема монетного обращения в Плиске (возрастания и убывания) за период 800-1120 г.

Выводы

Из проведенных выше анализов, из диаграммы и графиков на **рис. 2**, **рис. 3** и **рис. 4** вытекает, что с большой вероятностью Плиска функционировала как полноценный средневековый город в интервале с около 970 до около 1100 г.; в период с II до VI века и с начала IX века до около 970 г. на этой территории скорее всего существовало небольшое поселение, а из периода 600 – 800 г. нет следов от нормальной городской или деревенской жизни (хотя можно было бы допустить, что на этом месте был укрепленный военный лагерь, или крепость, или замок).

Этот вывод не соответствует принятому в исторической науке представлению о создании, развитии и функционировании города Плиска. Очевидно необходим новый, критический анализ первоисточников (в том числе и игнорируемых до сих пор), нужна проверка датировок монет и интерпретации археологических находок, чтобы найти причины установленных противоречий и устранить их.

Библиография

- [Каждан, 1954] Каждан, А. П. Византийские города в VII-XI вв. Советская археология, XXI, 1954, 164-188.
- [Соколова, 1958] Соколова, И. В. Клады византийских монет как источник для истории Византии VII-XI вв. Виз. Временник, т. XV, 1958, 50-63.
- [Gândilă, 2009] Gândilă, A. Early Byzantine Coin Circulation in the Eastern Provinces: A Comparative Statistical Approach. AJN Second Series 21 (2009) pp. 151–226.
- [Metcalf, 1958] Metcalf, D. M. Statistische Analyse bei der Auswertung von Munzfundmaterialien. Jahrbuch für Numismatik und Geldgeschichte 9 (1958), 187–196.
- [Metcalf, 1960] Metcalf, D.M. The currency of Byzantine Sirmia and Slavonia. Hamburger Beiträge zur Numismatik 14 (1960), 429–444.
- [Morrisson, 1980] Morrisson, C. Les monnaies. In: J.-P. Sodini, G. Tate, B. Bavant, S. Bavant, J.-L. Biscop, D. Orssaud, and C. Morrisson, Dehes (Aurie du nord) campagnes I–III (1976–1978), recherches sur l'habitat rural. Paris: Librairie Orientaliste Paul Geuthner, 1980. pp. 267–287.
- [Morrisson, 1981] Morrisson, C. Estimation du volume des émissions de Solidi de Tibère et Maurice à Carthage (578–602). In: C. Carcassonne and T. Hackens, eds. Statistique et numismatique: table ronde organisée par le Centre de mathématique sociale de l'École des hautes études en sciences sociales de Paris et le Séminaire de Numismatique Marcel Hoc de l'Université Catholique de Louvain, Paris, 17–19 sept. 1979, Strasbourg, Conseil de l'Europe, Assemblée parlementaire, 1981. pp. 267–284.

- [Newton, 2006] Newton, D. P. Found Coins as Indicators of Coins in Circulation: Testing Some Assumptions. *European Journal of Archaeology* 2006; 9; 211. <http://eja.sagepub.com>
- [Йорданов, 2000] Йорданов, Ив. Монеты и печати от Плиска (1899 – 1999). Плиска-Преслав, 8, 2000, 135-167.
- [Фоменко, 1981а] Фоменко, А. Новые экспериментально-статистические методики датирования древних событий и приложения к глобальной хронологии древнего мира. Препринт Гос. Ком. Телев. Радиовещ. 3672, № Б07201 (от 9/-81), Москва, 1981.
- [Фоменко, 1981б] Фоменко А. Т. Информативные функции и связанные с ними статистические закономерности. Тез. докл. 3-й Междунар. Вильнюс. конф. по теории вероятностей и мат. статистике. Вильнюс, 1981. Т. 2, 211-212.
- [Fomenko and Rachev, 1990] Fomenko, A. and S. Rachev. Volume Functions of Historical Texts and The Amplitude Correlation Principle. *Computers and the Humanities*, 11 (1990), 187-206.
- [Tabov, 2003] Tabov, J. Chronological Distribution of Information in Historical Texts. *Computers and the Humanities*, 24 (2003), 235-240.
- [Tabov, Vasilev and Velchev, 2003] Tabov, J., Vasilev K. and Velchev A. A mathematical model of monetary circulation in Medieval Bulgaria. *Storiadelmondo*. 2003: <http://www.storiadelmondo.com/14/tabov.monetary.pdf>
- [Tabov, J., A. Velchev, M. Dobрева, K. Sotirova, 2004] Tabov, J., A. Velchev, M. Dobрева, K. Sotirova. Chronological distribution of the Bulgarian mediaeval manuscripts preserved in Bulgaria. In: *Mathematics and Education in Mathematics. Proc. Of the Thirty Third Spring Conference of the Union of Bulgarian Mathematicians, Borovets, April 1-4, 2004*, 257-261.
- [Hristova and Dobрева, 2004] S. Hristova, M. Dobрева. Some observations on the chronological distribution of mediaeval manuscripts and church items preserved in Bulgaria. In: *Mathematics and Education in Mathematics. Proc. Of the Thirty Third Spring Conference of the Union of Bulgarian Mathematicians, Borovets, April 1-4, 2004*, 214-217.

Информация об авторах



Йордан Табов – Институт математики и информатики БАН, ул. Акад. Г. Бончев блок 8, 1113 София, Болгария; e-mail: tabov@math.bas.bg

Основные области научных исследований: Применения математики и информатики в гуманитарных науках, дидактика математики и информатики



Галина Панайотова – УНИБИТ – Университет библиотечного дела и информационных технологий, София, Болгария; e-mail: panayotovag@gmail.com

Основные области научных исследований: Гиперболические системы дифференциальных уравнений с частными производными, Математическое моделирование, Информационные технологии

ФОРМАЛИЗАЦИЯ МОДЕЛИ ЭВОЛЮЦИИ ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЕЙ

Галина Гайворонская

Аннотация: Предложено для описания телекоммуникационной сети во времени и пространстве использовать пять групп параметров: исследуемый период времени, топологическую структуру исследуемой сети, требования на обслуживание и увеличение емкости сети, технические и экономические параметры существующего и вводимого оборудования; экономические параметры эволюции сети. Для каждой группы параметров построены математические абстракции, формирующие общую пространственно-временную систему в которой может быть выполнено исследование эволюции ТС.

Ключевые слова: телекоммуникационная сеть, процесс введения принципиально нового оборудования математическая модель.

Ключевые слова классификации ACM: C.2. Computer-communication networks, H. Information Systems - H.1 Models and Principles, K. Computing Milieux - K.6 Management of computing and information system

Введение

В последние годы процесс развития телекоммуникационных сетей идет всевозрастающими темпами, при этом существенно изменилась общая ситуация в сфере инфокоммуникаций в связи с происходящим переходом от индустриального общества к информационному. Одновременно с этим изменились и общие подходы к созданию и развитию телекоммуникационных сетей (ТС). Это вызвано рядом причин, основными из которых являются увеличивающаяся конкуренция на рынке инфокоммуникационных услуг (ИКУ) и возрастающие темпы появления новых технологий в телекоммуникациях.

Бурный рост ТС происходит не только в связи с возрастающими потребностями в предоставлении постоянно увеличивающегося количества разнообразных ИКУ, но и благодаря невиданным ранее темпам смены сетевых технологий. Если раньше новые поколения систем коммутации и передачи появлялись на ТС раз в несколько десятилетий, то в настоящее время смена технологий происходит гораздо быстрее. Причем зачастую вводимые технологии оказывают значительное воздействие на структуру модернизируемой сети. При этом происходит не только количественный рост емкости и пропускной способности ТС, но и существенные качественные изменения, в большинстве случаев требующие пересмотра основных принципов построения сетей. Наглядные примеры: переход от аналоговой техники к цифровой, введение беспроводного доступа на стационарных сетях, переход от электрических сигналов к оптическим и введение на ТС оптических систем передачи и коммутации. В связи этим общие подходы к планированию ТС требуют в некоторых аспектах кардинальных изменений. Тщательная проработка принимаемых решений при средне- и долгосрочном планировании выходит на первый план потому, что ошибки планирования ТС ведут не только к существенным финансовым потерям, чем еще десятилетие назад, но и, что не менее важно, их последствия гораздо труднее устранить. Все это доказывает важность исследования процессов эволюции ТС, связанных не только с удовлетворением спроса населения и предприятий на ИКУ, но и с внедрением принципиально новых технологий, существенно влияющих как на структуру модернизируемых сетей, так и на необходимые для такой модернизации инвестиции. И особенно важной становится проблема наиболее целесообразного использования выделенных для этой цели инвестиций с учетом изменения реальной стоимости вложенных средств с течением времени. Указанная проблема требует разработки модели, имитирующей процесс эволюции ТС.

Для того, что бы точнее определиться с объектом исследования следует обратиться к классификации основоположников теории связи Г.Б. Давыдову, В.Н. Рогинскому и А.Д. Харкевичу, по учебникам которых, посвященным теории сетей связи, до сих пор учатся студенты связных специальностей. По этой классификации выделяется три вида ТС: сигнальные сети, сети без коммутации (с долговременными кроссовыми соединениями) и сети с коммутацией (оперативными соединениями, устанавливаемыми на время одного интерактивного обмена, например, телефонного разговора или на время передачи одного сообщения). В настоящей работе исследуется коммутируемая ТС при ее модернизации от некоторого исходного состояния в процессе динамического развития требующего изменения основных принципов построения этой сети, без ухудшения текущей работы сети по предоставлению ИКУ.

Вопросы выбора коммутационного оборудования и алгоритмов управления процессами установления соединений на коммутируемой сети выходят на первый план, поскольку в последние годы существенно изменилось соотношение затрат на отдельные элементы создаваемой или модернизируемой ТС. Если несколько десятилетий назад определяющими были затраты на линейные сооружения и каналобразующее оборудование, то впоследствии эта ситуация существенно изменилась в сторону увеличения доли затрат на коммутационное оборудование, и сетевое программное обеспечение [1]. Поэтому и формулировка задач сетевой оптимизации тоже претерпела кардинальные изменения. В настоящее время можно позволить себе создать первичную сеть с заранее заданной и значительной избыточностью, поскольку это несущественно увеличивает общую стоимость модернизируемой сети. Однако следует очень тщательно подходить к вопросам выбора типов и объемов коммутационного оборудования, поскольку доля затрат на эти системы является определяющей для создания и модернизации ТС [2].

Постановка задачи

Работа посвящена исследованию процесса эволюции ТС, при этом акцент в ней делается на переходный период, то есть переходный процесс. Известно, что именно переходные процессы представляют наибольший научный и практический интерес вне зависимости от того, где именно они происходят, в такой большой системе как ТС или просто в электрической цепи. И в этой связи важно правильно планировать переходный период, что бы не ухудшить качество работы сети и наиболее рационально истратить выделенные для этого средства. Следовательно, необходимо оптимизировать процесс эволюции таким образом, чтобы учесть всю совокупность качественных, экономических и временных показателей.

При этом исследуется эволюция сети от некоторого исходного состояния, что значительно усложняет задачу (создать что-либо заново гораздо легче, чем внести исправления в существующую структуру не прерывая и не ухудшая ее работы). Учет одновременно пространственного и временного аспектов эволюции и оптимизация динамики эволюции сети от заданного исходного состояния являются двумя важными аспектами задачи исследования. Третьим не менее важным аспектом является то, что исследуется процесс эволюции сети при введении оборудования, характеристики которого требуют пересмотра основных принципов построения сети. То есть понятие «принципиально новое оборудование» предполагает использование технологий, применение которых требует нового подхода к планированию ТС. При этом для реализации постепенного перехода от существующих технологий к новым, на сети должно быть обеспечено сосуществование, как старого, так и нового оборудования, что может быть реализовано методом наложения, либо замены на отдельных узлах исследуемой сети [3]. Поэтому необходимо проанализировать модель эволюции сети, использующую два вида оборудования, которые можно обозначить как существующее и «перспективное» или новое. Такой анализ может использоваться при введении любых новых технологий на действующей ТС [4].

Четвертой особенностью рассматриваемой постановки задачи являются предположения о типах нового оборудования. При этом в работе рассматривается два вида нового оборудования: основное или опорное и выносное или дистанционно управляемое. То есть вторым типом устанавливаемого оборудования

являются системы, подключаемые к оборудованию первого типа и функционирующие с ним взаимосвязано, в то время как первый тип оборудования может работать как автономно, так и вместе с оборудованием второго типа. Примеров оборудования, взаимодействующего таким образом множество. Это и цифровые системы коммутации с подключаемыми к ним выносными модулями (ВМ) – цифровыми концентраторами; и центры коммутации сотовых сетей мобильной связи с контроллерами базовых станций, предназначенные для концентрации нагрузки от группы БС и распределенные в пространстве модули гибких программируемых коммутаторов *Softswitch*.

Для ТС процесс введения принципиально нового оборудования не может быть одномоментным в силу как технических, так и экономических причин. Некоторые из этих причин проанализированы автором в работах [4-13]. Это сложная проблема, включающая в себя большой комплекс системных вопросов, обеспечивающий взаимодействие существующего и вновь вводимого оборудования. Учитывая консервативность ТС и необходимость ее непрерывной работы, очевидно, что процесс перехода от существующего оборудования к принципиально новому, будет скорее эволюционным, чем революционным и может растянуться на несколько лет. Длительность переходного периода зависит от многих факторов. В частности от [7]: состояния существующей сети, объемов инвестиций, сроков окупаемости существующего и вводимого оборудования, степени амортизации существующего оборудования, необходимой скорости роста существующей сети, определяемой, в первую очередь, требованиями пользователей на обслуживание сетью, а также доступности вновь вводимого оборудования, коэффициента инфляции и многого другого.

Одним из используемых подходов для решения задач оптимизации ТС является независимое рассмотрение отдельных узлов сети с учетом фактора времени [10]. Применение такого подхода не позволяет в общем случае исследовать процесс развития сети, поскольку не учитываются пространственные связи. Другим подходом является рассмотрение сети в определенный момент времени, учитывая лишь пространственные связи и не принимая в расчет временных.

В связи с разработкой многоуровневой модели сети можно утверждать, что средства анализа стационарных сетей в целом созданы. Исследования же динамики сетей находятся на начальной стадии, так как традиционно используемый для исследования математический аппарат (теория массового обслуживания, теория надежности), да и не традиционный аппарат (теория конечных полумарковских процессов, теория конечных цепей Маркова и др.) по своей природе является средством описания стационарных систем. Между тем реальные сети являются нестационарными объектами. Их нестационарность определяется, в частности, сезонными и суточными изменениями нагрузки, перемещениями пользователей сети, выходом из строя и восстановлением сетевых элементов, реконфигурацией сети и др. Отсутствие адекватного, по возможности несложного аппарата исследования нестационарных сетей приводит к тому, что сети рассчитываются на максимальные, пиковые нагрузки, что существенно увеличивает затраты на создание и модернизацию ТС.

Исходя из изложенного, задачу, исследуемую в работе можно сформулировать следующим образом: разработка модели, позволяющей оптимизировать процесс динамического изменения ТС от существующего состояния при введении двух видов взаимосвязанного оборудования, параметры которого требуют принципиального изменения характеристик исследуемой сети с учетом пространственных и временных факторов. Исследуемая задача не нова, однако в такой постановке она предлагается впервые.

Очевидно, что в своей общей постановке проблема оптимального синтеза ТС многокритериальна и требует правила, позволяющего сравнивать возможные альтернативы. В простейшем случае такое правило может быть задано скалярной функцией на множестве возможных вариантов, а наилучшее решение определяется из условий экстремума этой функции. Однако в практических задачах построение такой функции вызывает серьезные затруднения. К тому же формирование целевых функций на начальных этапах выбора приводит к максимальному субъективизму и заранее запрограммированному результату, исключая широкий поиск возможных кандидатов на оптимальность по менее сильным

критериям. Процедуры выбора дают возможность выявить рациональные решения и позволяют разрешить компромиссы с учетом нескольких показателей. С другой стороны, всякое проектирование многовариантно по своему существу, а это требует многоаспектной, многокритериальной оценки возможных вариантов проектов с целью выбора наиболее целесообразного решения. Если рассматривать наиболее часто встречающуюся триаду критериев: качество, время, стоимость, то в нашем случае, совокупность показателей качества предоставления услуг, надежности и эффективности ТС введены в виде ограничений, а оптимизируются время, то есть момент введения оборудования, и стоимость модернизации сети, как на каждом отдельном этапе проектирования, так и суммарная за весь период исследования. Выбор экономических критериев, а именно действительной стоимости результирующей сети и действительных годовых затрат существенно усложняет задачу, так как требует введения дисконтного коэффициента, который вносит экспоненциальную составляющую в функцию стоимости развития сети. В результате этого задача становится нелинейной.

Общая структура решения задачи

Общеизвестным подходом для решения комплексных проблем является декомпозиция общей проблемы на ряд отдельных задач последовательно решаемых в процессе исследований. Первая задача, которой посвящена данная статья, предполагает введение ряда математических абстракций, позволяющих формализовать проблему. Для этого, прежде всего, выделим используемые исходные параметры и представим задачу в обобщенном виде как показано на рис. 1

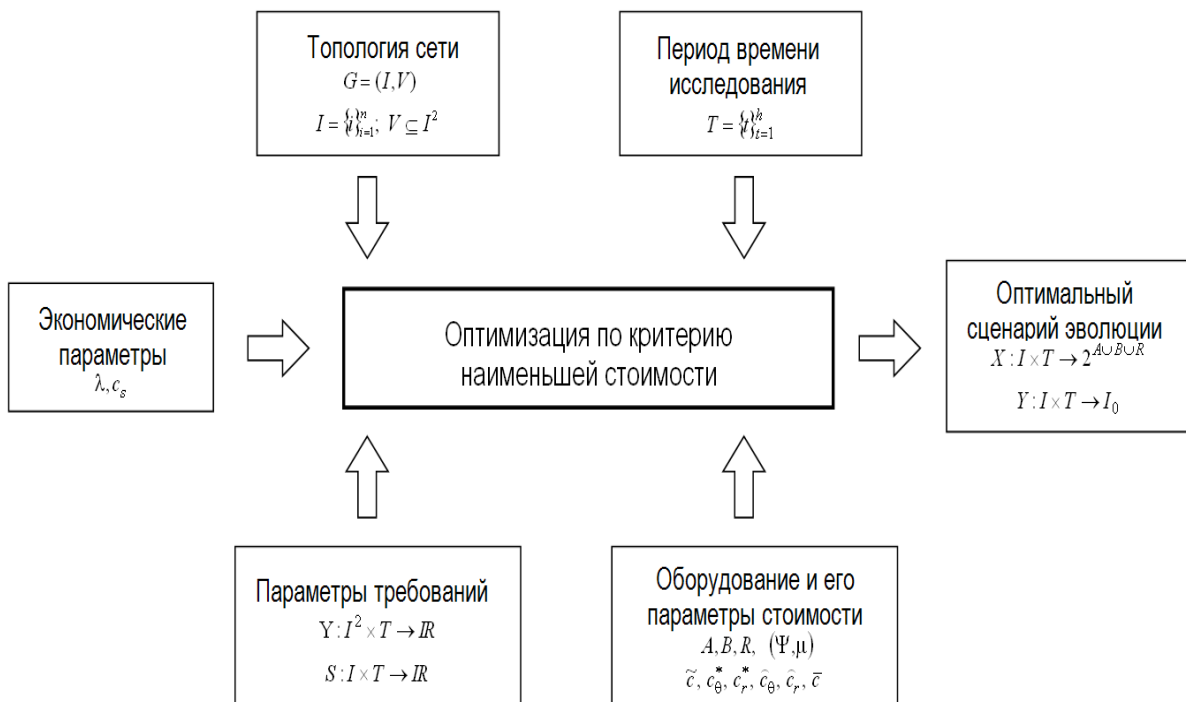


Рис.1. Модель оптимизации телекоммуникационной сети

Для каждого входного параметра, используемого в модели пространственно-временной структуры ТС предложено формализованное описание, при этом входные параметры модели сгруппированы. Выделены следующие группы параметров, описывающих сеть в пространстве и времени [14]:

- топологическая структура сети, то есть рассматриваемое пространство;
- исследуемый период времени;
- требования на обслуживание пользователей и на развитие сети;

- параметры доступного оборудования;
- экономические параметры.

Для каждой группы параметров построен математический аппарат точно, обще, полно и удобно описывающий реальные объекты, в контексте поставленной задачи, то есть формализовано представление исходных данных, используемых при планировании сетей.

Топологическая структура сети. Обозначим через I множество всех узлов коммутации (УК), как существующих на момент начала исследования, так и потенциальных, которые могут быть введены в течение исследуемого периода, а через n – количество таких объектов, которые в дальнейшем будем называть узлами сети. Другими словами, мощность множества узлов есть $n \ ||I| = n$. Рассмотрение в качестве узлов сети объектов, соответствующих узлам, которые, могут быть введены в течение периода исследования (потенциальных точек размещения узлов), позволяет сделать рассматриваемое пространство инвариантным во времени. При этом построен аппарат, обеспечивающий возможность исключить влияние тех узлов сети, которые в рассматриваемый момент времени еще не были введены. Природа объектов из множества I для нас не существенна, поэтому их можно считать последовательными натуральными числами, что определяет I следующим образом $I = \{i \in \mathbb{N} \mid 1 \leq i \leq n\} = \{1, 2, \dots, n\}$. Тогда, любые интересующие нас характеристики узлов можно рассматривать как функции, определенные на множестве узлов.

Пусть теперь V – множество пучков каналов между узлами сети. Предположим, что каждую пару узлов соединяет не более одного пучка каналов. В связи с чем, пучок каналов можно рассматривать как упорядоченную пару узлов, первый элемент которой определяет узел, из которого выходит пучок, а второй – узел в который он входит. То есть $V \subseteq I^2$, где I^2 – декартов квадрат множества I или множество пар данного множества. Аналогично, дополнительные характеристики пучков будем рассматривать в виде функций, определенных на множестве V . Кроме того, можно считать, что пары вида $\langle i, i \rangle \ i \in I$ не имеют смысла как пучки каналов. Иначе говоря $\forall i \in I \ \langle i, i \rangle \notin V$.

В соответствии с указанными свойствами, система полученных множеств I и V образует конечный ориентированный граф G без петель и кратных ребер $G=(I, V)$. В этом графе I составляет множество вершин графа, а V – множество ребер. Функция Δ должна удовлетворять ряду условий:

1. быть неотрицательной в своей области определения $\forall i, j \in I \ \Delta(i, j) \geq 0$;
2. для пары узлов, в которой первый элемент совпадает со вторым значение функции должно быть нулевым $\forall i \in I \ \Delta(i, i) = 0$;
3. симметричность Δ относительно своих параметров, то есть рассматриваемая характеристика должна быть одинаковой как для пары $\langle i, j \rangle \ i \in I^2$, так и для пары $\langle j, i \rangle \ \forall i, j \in I \ \Delta(i, j) = \Delta(j, i)$;
4. для функции должно выполняться неравенство треугольника, а именно прямое расстояние между любой парой узлов не может превышать расстояния через некоторый промежуточный узел, то есть $\forall i, j, k \in I \ \Delta(i, j) \leq \Delta(i, k) + \Delta(k, j)$;

Сформулированные условия гарантируют, что функция Δ является метрикой на множестве узлов сети.

Пусть D – квадратная действительная матрица порядка n над действительным полем $D = \|\|d_{ij}\|\| \in M_n(\mathbb{R})$. Определим элементы этой матрицы как $d_{ij} = \Delta(i, j)$. То есть D – матрица расстояний между узлами, имеющая вид

$$D = \begin{pmatrix} 0 & d_{12} & d_{13} & \dots & d_{1n} \\ d_{12} & 0 & d_{23} & \dots & d_{2n} \\ d_{13} & d_{23} & 0 & \dots & d_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{1n} & d_{2n} & d_{3n} & \dots & 0 \end{pmatrix}.$$

Из свойств Δ следует, что D является симметричной матрицей и содержит только нули на главной диагонали.

Исследуемый период времени. Планирование эволюции сети осуществляется для ограниченного промежутка времени. То есть, если t – некоторый момент исследуемого периода, то t удовлетворяет соотношению $\underline{t} \leq t \leq \bar{t}$, где \underline{t} – момент начала исследования; \bar{t} – момент конца исследования, $\underline{t} \neq \bar{t}$. При этом следует учитывать, что предоставление оборудования, расчет нагрузки и принятие решений может осуществляться лишь в некоторые моменты исследуемого периода. То есть, множество таких моментов времени образует некоторое разбиение действительного сегмента $[\underline{t}, \bar{t}]$ $t \in \tau_{[\underline{t}, \bar{t}]}$,

$$\tau_{[\underline{t}, \bar{t}]} = \{t_k\}_{k=0}^h : \underline{t} = t_0 < t_1 < t_2 < \dots < t_h = \bar{t} \quad (1)$$

где $\tau_{[\underline{t}, \bar{t}]}$ – разбиение сегмента $[\underline{t}, \bar{t}]$;

h – количество точек разбиения;

$t_k, k = \overline{0, h}$ – точки разбиения, соответствующие рассматриваемым моментам времени.

Промежутки между такими моментами могут быть, вообще говоря, различными, однако для упрощения дальнейших выкладок, будем считать их равно удаленными

$$\forall k, l = \overline{1, h} \quad t_k - t_{k-1} = t_l - t_{l-1} = \frac{\bar{t} - \underline{t}}{h} = \text{const} \quad (2)$$

То есть последовательность $\{t_k\}_{k=0}^h$ является арифметической прогрессией

$$\forall k = \overline{1, h} \quad t_k = \underline{t} + (k-1)\Delta t$$

где $\Delta t = \frac{\bar{t} - \underline{t}}{h}$ – разность прогрессии.

Это предположение не ограничивает общности, как видно из доказательства следующей леммы.

Лемма 1. Для любого сегмента $[\underline{t}, \bar{t}]$ и его разбиения $\tau_{[\underline{t}, \bar{t}]}$ можно построить эквивалентное с точки зрения решаемой задачи разбиение $\tau_{[\underline{t}^*, \bar{t}^*]}$ некоторого другого сегмента $[\underline{t}^*, \bar{t}^*]$, удовлетворяющее (2).

Доказательство. Предположим, что все точки t_k являются рациональными числами. Если это не так, то есть существует некоторая точка t_k , являющаяся иррациональной, то можно выбрать целое число, сколь угодно близкое к t_k $\forall \varepsilon > 0 \exists t'_k \in \mathbb{Q} : |t'_k - t_k| < \varepsilon$.

Так как, в любом случае исходные измерения обладают определенной погрешностью, и результат

требуется лишь с некоторой точностью, всегда можно заменить иррациональное t_k соответствующе подобранным рациональным t'_k , так, чтобы это не повлияло на точность моделирования. Итак, все рассматриваемые точки t_k являются рациональными, то есть, представимы в виде дробей

$$t_k = \frac{m_k}{n_k}, m_k \in \mathbb{Z}, n_k \in \mathbb{N}.$$

Определим \hat{n} как наименьшее общее кратное всех знаменателей n_k и построим новое разбиение, точки \hat{t}_k которого определяются так $\hat{t}_k = \hat{n}t_k$. Учитывая то, что множество точек t_k ограничено и образует разбиение, и не отрицательность \hat{n} имеем $\hat{t} \equiv \hat{t}_1 < \hat{t}_2 < \dots < \hat{t}_h \equiv \hat{t}$. Так как \hat{n} делится на любое n_k $\forall k = \overline{1, h}$ $\hat{n} : n_k$, все полученные новые точки \hat{t}_k будут целыми числами, то есть $\forall k = \overline{1, h}$ $\hat{t}_k \in \mathbb{Z}$. Дополним точки \hat{t}_k оставшимися целыми числами из сегмента $[\hat{t}, \hat{t}]$, множество которых обозначим $\tilde{\tau}$. Условимся, что сеть не меняется в точках $\tilde{\tau}$. Обозначим полученное разбиение $\tau_{[\hat{t}^*, \hat{t}^*]}^*$

$$\tau_{[\hat{t}^*, \hat{t}^*]}^* : \hat{t}^* = t_1^* < t_2^* < \dots < t_{h^*}^* = \bar{t}^* \tag{3}$$

Тогда для каждой точки t_k исходного разбиения $\tau_{[\underline{t}, \bar{t}]}$ в полученном разбиении существует единственная соответствующая ей точка $t_{k^*}^*$. Учитывая нумерацию в (3), k^* может не совпадать с k , поэтому определим дополнительную функцию, отображающую множество натуральных чисел в себя же $\eta : \mathbb{N} \rightarrow \mathbb{N}$ такую, что $\eta(k) = k^*$.

Теперь сформулируем важное свойство полученного разбиения $\tau_{[\hat{t}^*, \hat{t}^*]}^*$: расстояние между любыми двумя точками $\tau_{[\underline{t}, \bar{t}]}$ совпадает с расстоянием между соответствующими точками $\tau_{[\hat{t}^*, \hat{t}^*]}^*$ с точностью до фиксированного коэффициента масштабирования. Это означает что,

$$\forall t_k, t_l \in \tau_{[\underline{t}, \bar{t}]} \quad \exists! t_{\eta(k)}^*, t_{\eta(l)}^* \in \tau_{[\hat{t}^*, \hat{t}^*]}^* : |t_k - t_l| = |t_{\eta(k)}^* - t_{\eta(l)}^*| \hat{n} \tag{4}$$

Этот факт, в сочетании с соглашением о неизменности сети в дополнительных точках $\tilde{\tau}$, обеспечивает неизменность получаемых результатов при замене $\tau_{[\underline{t}, \bar{t}]}$ на $\tau_{[\hat{t}^*, \hat{t}^*]}^*$, с точностью до масштабирования шкалы времени. Что и завершает доказательство леммы. ■

Не смотря на конструктивность приведенного доказательства, использованная процедура, в силу своей значительной избыточности, не вполне подходит для применения на практике. Тем не менее, доказательство показывает возможность упрощения сложных математических структур, используемых при описании сети, без уменьшения общности и потери точности. Следует отметить, что не вполне формальный характер некоторых рассуждений связан лишь с упрощением выкладок, не несущих значительной нагрузки в масштабе решаемой задачи. Продолжим упрощение арифметической прогрессии (1), положив $\underline{t} = 0$. Это вполне обоснованно, ввиду того, что для модели не важны реальные даты. Тогда

$\bar{t} = h$ и последовательность $\{t_k\}_{k=0}^h$ примет вид $\{0, 1, 2, \dots, h\}$.

Обозначим полученное через T ,

$$T = \{k\}_{k=0}^h = \{0, 1, 2, \dots, h\}.$$

Таким образом, T будет являться моделью исследуемого периода времени. Обозначим через $\theta, r \in T$ – соответственно моменты наложения и замены оборудования существующих узлов коммутации (УК). Процесс развития сети по отношению к переменным θ, r изображен на рис. 2 [12]. Поскольку замена осуществляется после наложения, применяется ограничение

$$0 \leq \theta < r \leq h. \tag{5}$$

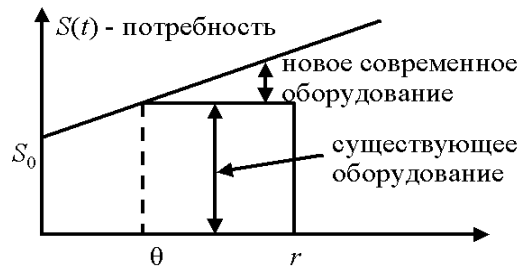


Рис. 2. Процесс эволюции сети

Требования на обслуживание. Определив рассматриваемое пространство в виде графа G , а время в виде последовательности T , перейдем к рассмотрению характеристик сети, существующих в этой системе. Обозначим через Y нагрузку между узлами сети и рассмотрим требования на ее обслуживание. Очевидно, эта величина зависит от рассматриваемой пары узлов и времени. Тогда, нагрузка представляет функцию $Y : I^2 \times T \rightarrow IR$. Значения этой функции, иллюстрируемой рис. 3 могут соответствовать как величинам реальных измерений между узлами в заданное время, так и прогнозируемым параметрам.

Рассмотрим пространственное сечение междуузловой нагрузки. Зафиксируем пару узлов $\langle i, j \rangle$ и

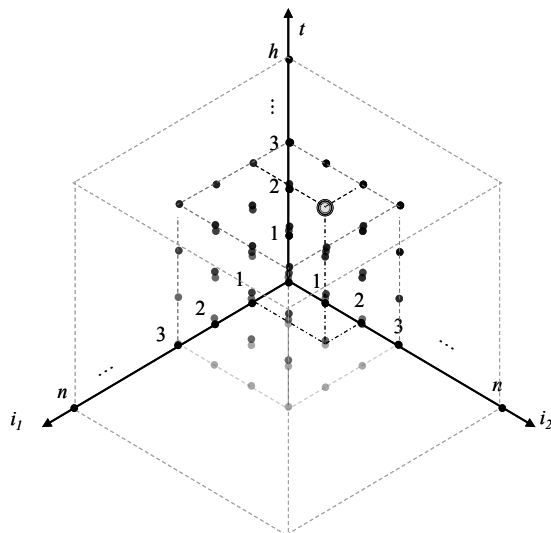


Рис. 3. Пространственно-временное представление междуузловой нагрузки

определим ее пространственное сечение \mathfrak{Y}_{ij} как

$$\mathfrak{Y}_{ij} : T \rightarrow IR, \quad \mathfrak{Y}_{ij}(t) = Y(i, j, t).$$

Таким образом, для каждой пары узлов определена функция одного аргумента, определяющая нагрузку между этой парой узлов в каждый момент времени. Из этой системы несложно получить временной срез нагрузки. Для этого зафиксируем некоторый момент времени t и построим квадратную матрицу n -ого порядка $Y(t)$ так, чтобы $Y(t) = \|\mathfrak{Y}_{ij}(t)\| \in M_n(IR)$.

Учитывая дискретность и ограниченность T , можно говорить о Y как о конечной системе матриц нагрузки.

Кроме требований на соединения между пользователями, сеть также может получать требования на подключение новых пользователей. Количество таких требований

будем рассматривать не для всей сети целиком, а для каждого узла в отдельности, то есть уже после

того, как запросы будут обработаны сетевым оператором и распределены по узлам. Поступление требований пользователей на подключение к сети, является стохастическим процессом, однако, в соответствии с анализом, проведенным автором в [18], будем рассматривать его как детерминированную функцию, однозначно распределенную в пространстве и времени.

Определим функцию $S : I \times T \rightarrow IR$, задающую требования $S(i, t)$ на подключение к узлу i в момент t . Зафиксировав произвольный узел, получим систему функций одного аргумента $s_i : T \rightarrow IR$, каждая функция, которой определяет требования на подключение для отдельного узла. Функции Y и S , а также их срезы определяют параметры требований на подключение к сети новых пользователей.

Оборудование и его стоимостные параметры. В соответствии с поставленной задачей, выделим три класса типов оборудования (рис. 4), используемого на сети: оборудование существующей технологии; опорное оборудование новой технологии; выносное оборудование новой технологии.

Обозначим эти классы оборудования в виде множеств A , B и R . Естественно считать эти множества конечными, каждый элемент которых a_i , b_i или r_i представляет тип оборудования соответствующего класса. Теперь определим стоимостные параметры, для оборудования каждого типа в виде следующих функций [8]:



Рис. 4. Классификация типов оборудования

$\tilde{c} : A \cup B \cup R \rightarrow IR$ – годовые расходы на обслуживание единицы оборудования некоторого типа;

$c_{\theta}^* : B \cup R \rightarrow IR$ – расходы на запуск нового оборудования при наложении;

$c_r^* : B \cup R \rightarrow IR$ – расходы на запуск нового оборудования при замене;

$\hat{c} : A \rightarrow IR$ – расходы на единичное приращение емкости существующего оборудования;

$\hat{c}_{\theta} : B \cup R \rightarrow IR$ – расходы на единичное приращение емкости нового оборудования при наложении;

$\hat{c}_r : B \cup R \rightarrow IR$ – расходы на единичное приращение емкости нового оборудования при замене;

$\bar{c} : A \cup B \rightarrow IR$ – стоимость единицы длины кабеля.

Можно для упрощения обозначений вместо функций \hat{c}_{θ} и \hat{c}_r рассматривать их расширения, определенные на $A \cup B$ и на A , и действующие аналогично \hat{c} .

Экономические характеристики. При решении задачи рассчитываются все денежные потоки, включающие затраты на установку нового опорного и выносного оборудования, его обслуживание, амортизацию, налоги, затраты на аренду земли, здания, коммунальные платежи и т. п. В соответствии с выбранным критерием оптимизации для любых затрат, которые будут иметь место на протяжении исследуемого периода, необходимо рассчитать их действительную величину. Для этого потребуется использовать дисконтную ставку i , то есть учесть процент инфляции [15]. В расчетах для приведения будущих затрат к текущему периоду, применяем дисконтный коэффициент, определяемый как $\frac{1}{(1+i)^t}$,

где i – дисконтная ставка, t – момент времени, соответствующий будущим величинам затрат. Для

удобства дальнейших рассуждений преобразуем традиционное представление дисконтного

$$\text{коэффициента, следующим образом } C^* = \frac{C_t}{(1+i)^t} = \left[\begin{array}{l} \frac{1}{(1+i)^t} = e^{-\rho t} \\ \rho = \ln(1+i) \end{array} \right] = C_t e^{-\rho t},$$

где C^* – действительная стоимость будущих затрат;

C_t – величина будущих затрат;

i – дисконтная ставка;

ρ – модифицированное представление дисконтного коэффициента;

t – момент времени, соответствующий периоду, в который рассматриваются будущие затраты.

Для учета возможности повторного использования оборудования, выводимого из эксплуатации, введем чистую остаточную стоимость единицы существующего оборудования [16] $c_s : A \rightarrow \mathbb{R}$.

Функционирование модели стратегии эволюции сети

Ряд исходных параметров модели вводится в виде ограничений, в частности, параметры существующего и вновь вводимого оборудования. Эти параметры также необходимо формализовать, что выполнено автором в работе [17]. В первую очередь, определим тип оборудования, устанавливаемого на узлах сети в каждый момент времени, в виде отображения $X : I \times T \rightarrow 2^{A \cup B \cup R}$, где $2^{A \cup B \cup R}$ – булеан множества $A \cup B \cup R$, то есть множество всех подмножеств множества типов оборудования. Такое задание X означает, что в каждый момент времени t , на некотором узле i может быть одновременно установлено несколько типов оборудования $X(i, t) \subseteq A \cup B \cup R$. Введем дополнительные обозначения этой характеристики, для этого зафиксируем некоторый узел i и определим

$$x_i : T \rightarrow 2^{A \cup B \cup R} \quad (6)$$

В силу технических особенностей, функции (5) не могут вести себя произвольно, то есть существуют некоторые ограничения на изменение типов установленного на узле оборудования.

Назовем множество типов оборудования, используемых в данный момент на узле, и соответственно, функции, выполняемые узлом – состоянием узла. Тогда $\Psi = 2^{A \cup B \cup R}$ – множество состояний узла. Зададим на множестве Ψ бинарное отношение μ , определяющее возможность перехода между двумя состояниями узла. Здесь под бинарным отношением понимается некоторая взаимосвязь между двумя элементами множества, которая либо имеет место, либо не имеет. Строго говоря, $(\langle \psi_1, \psi_2 \rangle \in \mu) \equiv (\psi_1 \mu \psi_2) \Leftrightarrow$ переход из ψ_1 в ψ_2 возможен. Следует отметить, что не во всякий момент времени, узел совершает переход в новое состояние – он может оставаться в прежнем. Таким образом, можно сказать, что допустим переход из состояния ψ_1 в это же состояние ψ_1 ,

$$\psi_1 \mu \psi_1 \quad \forall \psi_1 \in \Psi \quad (7)$$

Так как речь идет о модернизации узла, с переходом от старых технологий к новым, принимается предположение о необратимости эволюции. Это предположение заключается в том, что если на некотором узле установлено оборудование k -поколения, оно не может быть заменено оборудованием более ранних поколений. Таким образом переход из состояния ψ_1 в состояние ψ_2 допустим, а обратный переход из состояния ψ_2 в ψ_1 невозможен, за исключением случая, когда $\psi_1 = \psi_2$. То есть

$$\left. \begin{array}{l} \psi_1 \mu \psi_2 \\ \psi_1 \neq \psi_2 \end{array} \right\} \Rightarrow \psi_2 \bar{\mu} \psi_1 \quad \forall \psi_1, \psi_2 \in \Psi .$$

Иначе говоря

$$\left. \begin{array}{l} \psi_1 \mu \psi_2 \\ \psi_2 \mu \psi_1 \end{array} \right\} \Rightarrow \psi_1 = \psi_2 \quad \forall \psi_1, \psi_2 \in \Psi \quad (8)$$

То есть отношение μ является антисимметричным.

Кроме того, переход между состояниями может осуществляться не непосредственно, а через некоторое промежуточное состояние. Такой переход возможен тогда и только тогда, когда возможны переходы из исходного состояния в промежуточное и из промежуточного в конечное. Последнее означает, что отношение μ обладает свойством транзитивности:

$$\psi_1 \mu \psi_2 \wedge \psi_2 \mu \psi_3 \Rightarrow \psi_1 \mu \psi_3 \quad \forall \psi_1, \psi_2, \psi_3 \in \Psi \quad (9)$$

Из показанных свойств (7), (8) и (9) можно сделать вывод о том, что отношение μ является отношением частичного порядка. Следовательно, система (Ψ, μ) образует частично упорядоченное множество (ЧУМ).

Для узлов, на которых установлено выносное оборудование, необходимо определить опорный узел, к которому подключен конкретный ВМ в данный момент времени. Для этого зададим отображение

$$Y : E_R \rightarrow I_B, \quad (10)$$

где $E_R = \{ \langle i, t \rangle \mid i \in I, t \in T : X(i, t) \cap R \neq \emptyset \}$ – множество точек пространственно-временной системы, требующих опорного оборудования, то есть упорядоченных пар из момента времени и узла, на котором расположен ВМ.

$I_B = \{ i \mid i \in I : \exists t \in T X(i, t) \cap B \neq \emptyset \}$ – множество потенциально опорных точек пространства, то есть узлов, на которых в течение исследуемого периода при замене введено опорное оборудование.

Это отображение в своей области определения задает опорный узел $Y(i, t) \in I$, к которому подключен узел i в момент t . Построение множеств E_R и I_B является довольно сложным и не вполне удобным в использовании, поэтому будем рассматривать также расширение функции (9) $Y : I \times T \rightarrow I_0$, где $I_0 = I \cup \{0\}$ – множество узлов сети, заполненное нулем. При этом предполагается, что на все x точках $(I \times T) \setminus E_R$ это расширение принимает значение нуль.

Введем также пространственные срезы для заданной пространственно-временной функции $y_i(t) = Y(i, t) \quad \forall i \in I$.

Исследование задачи в пространственно-временной системе

Учитывая приведенные выше построения, решение поставленной задачи можно рассмотреть в виде пары (X, Y) , где $X : I \times T \rightarrow 2^{A \cup B \cup R}$ определяет используемые типы оборудования на узлах, а $Y : I \rightarrow I_0$ – структуру подключения выносов. Сформулируем это утверждение в виде леммы.

Лемма 2. На основании пары (X, Y) всегда можно определить соответствующую стратегию эволюции сети и при том только одну.

Доказательство. В соответствии с соглашением о допустимых стратегиях эволюции сети, введение нового типа оборудования на узле может сопровождаться либо полной заменой оборудования предыдущего типа оборудованием нового типа, либо сохранением оборудования предыдущего типа и использованием оборудования нового типа исключительно для расширения емкости узла. Кроме того, в один момент времени допускается введение только одного типа оборудования на одном узле.

Из вышеизложенного следует, что если в момент времени t на узле i происходит изменение набора

используемых типов оборудования $X(i, t) \neq X(i, t-1)$, то возможны только два варианта $X(i, t) = X(i, t-1) \cup \{\chi_2\}$ или $X(i, t) = (X(i, t-1) \setminus \{\chi_1\}) \cup \{\chi_2\}$ где $\chi_2 \in B \cup R$ – тип оборудования, вводимый в момент t ; $\chi_1 \in B \cup R$ – тип оборудования, введенный непосредственно перед χ_2 .

Первый вариант соответствует наложению оборудования типа χ_2 на существующее оборудование, а второй – полной замене оборудования типа χ_1 на оборудование типа χ_2 . Эти варианты для каждого узла i и момента времени t определяются однозначно и в свою очередь однозначно определяют стратегию развития, что и требовалось доказать. ■

При использовании линейной функции, требования на изменение емкости сети имеют вид $S(t) = s_0 + gt$, где s_0 – требования на подключение к узлу в начальный момент времени; g – скорость увеличения емкости узла.

Пусть $w \geq s_0$ – монтируемая емкость сети в момент $t = 0$ и $e = \frac{w - s_0}{g}$ – первое время исчерпания

ресурсов, определяемое как: $e = \left\lfloor \frac{w - s_0}{g} \right\rfloor$. В соответствии с введенными обозначениями, $\theta, r \in T$ –

соответственно моменты наложения и замены оборудования, при этом $0 \leq \theta < r \leq h$.

Принятые определения иллюстрирует рис. 5. Из рисунка видно, что при линейном росте требований, емкость узла наращивается ступенчато, в соответствии с допущением о дискретности промежутков принятия решений. При этом первая ступень может быть шире последующих, что обуславливается возможной избыточной емкостью сети на момент начала исследования.

Случай, показанный на рис. 5 соответствует наличию на узле оборудования в начальный момент времени. В течение периода $[0, \theta]$ происходит увеличение емкости узла оборудованием того же типа. Момент θ соответствует наложению нового оборудования. Далее увеличение емкости узла происходит добавлением оборудования именно этого типа, вплоть до момента r . В момент r происходит замена всего установленного существовавшего оборудования на оборудование новой технологии.

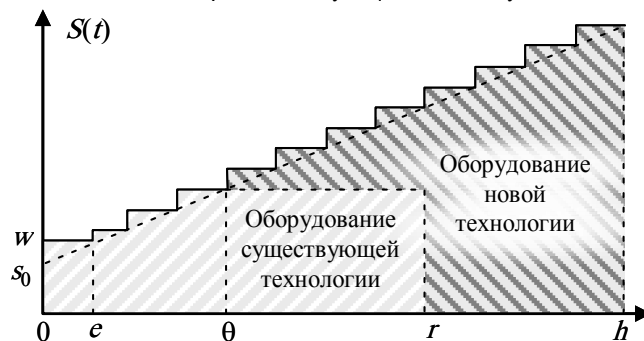


Рис. 5 Изменение емкости узла во времени

Заключение

Таким образом в работе предложено для описания сети во времени и пространстве использовать пять групп параметров: исследуемый период времени, топологическую структуру исследуемой сети, требования на обслуживание и увеличение емкости сети, технические и экономические параметры существующего и вводимого оборудования; экономические параметры эволюции сети. Для каждой группы параметров построены математические абстракции, формирующие общую пространственно-временную систему в которой выполнено исследование эволюции ТС. Показано, что пучок каналов можно рассматривать как упорядоченную пару узлов, первый элемент которой определяет узел, из которого выходит пучок, а второй – узел в который он входит, а расстояние между этими узлами как функцию $\Delta: I^2 \rightarrow \mathbf{R}$, неотрицательную в своей области определения, симметричную относительно ее параметров и отображающую пары узлов во множество вещественных чисел. Сформулированы условия, которым должна отвечать предложенная функция для того, чтобы являться метрикой на множестве узлов

сети и описывать не только физическое расстояние между соответствующими узлами, но и иные характеристики, например, длину прокладки кабеля между этими узлами. Доказано, что исследуемый период времени можно представить в виде множества рациональных равноудаленных точек действительного сегмента. Выполненное доказательство показывает возможность упрощения сложных математических структур, используемых при описании сети, без уменьшения общности и потери точности и позволяет построить модель исследуемого периода времени для аналитической модели исследования пространственно-временной структуры эволюции ТС. Межузловая нагрузка в каждый момент времени представлена в виде числовой функции $Y: I^2 \times T \rightarrow \mathbb{R}$, заданной на декартовом произведении множества пар узлов и рассматриваемых моментов времени. Значения этой функции могут соответствовать как величинам реальных измерений между узлами в заданное время, так и полученным в результате прогноза. Учитывая дискретность и ограниченность рассматриваемого временного промежутка, получена конечная система матриц нагрузки, позволяющая реализовать пространственно-временное представление межузловой нагрузки необходимое для аналитического исследования эволюции ТС во времени и пространстве.

Впервые введены формализованные описания типов коммутационного оборудования устанавливаемого на узлах сети в каждый момент времени в виде отображения $X: I \times T \rightarrow 2^{A \cup B \cup R}$ и стоимостных параметров оборудования, необходимые для создания аналитической модели исследования пространственно-временной структуры эволюции ТС. Предложено модифицированное представление дисконтного коэффициента для приведения значений будущих денежных потоков к текущему периоду. Предложено рассматривать состояния узла и допустимые переходы между ними в виде частично упорядоченного множества, что позволило формализовать описание любого из возможных состояний узла и любое количество переходов. А для задания опорного узла, к которому подключен конкретный выносной модуль в конкретный момент времени, предложено использовать отображение вида $Y: E_R \rightarrow I_B$, в котором E_R представляет собой множество точек пространственно-временной системы требующих опорного оборудования, а I_B – множество точек пространства, на которых в течение исследуемого периода введено опорное оборудование. Построенное отображение дает возможность формализовать структуру подключения выносных модулей к опорным узлам в модели исследования пространственно-временной структуры эволюции ТС. Доказано, что абстрактная модель процесса эволюции сети может быть представлена в виде пары функций (X, Y) , описывающих типы оборудования, установленного на каждом узле в каждый момент времени и структуру подключения выносных модулей.

Основная сложность задачи обусловлена ее пространственно-временной структурой. Именно необходимость одновременного учета изменений сети, как в пространстве, так и во времени, определяет комплексность решаемой задачи, следовательно, для ее решения необходим комплексный подход, сочетающий аналитические и алгоритмические методы. Однако, при введении некоторых ограничений на время или пространство, задача упрощается и допускает полностью аналитическое решение. В частности при рассмотрении только пространственной задачи в фиксированный момент времени или только временной задачи в одной точке пространства, решение становится как минимум на порядок проще. Ограничения на рассматриваемое пространство – в данном случае топологическую структуру сети, за исключением наиболее естественных, существенно сужают применение результатов исследования и не могут использоваться при исследовании проблемы в целом. Тем не менее, решение частных задач может оказаться весьма полезным при решении общей проблемы. Введение ограничений на время, может в ряде случаев привести к корректному и при этом более простому решению общей проблемы. Дальнейшие исследования позволили предложить обобщенное решение задачи аналитическими методами и построить имитационную модель процесса эволюции ТС, позволившую автоматизировать процесс получения оптимального сценария развития сети во времени и пространстве по критерию стоимостных параметров.

Благодарности

The paper is partially financed by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA and the Consortium FOI Bulgaria. www.ithea.org , www.foibg.com

Литература

- Соколов Н. А. Эволюция местных телефонных сетей ТОО. – Пермь: Книга, 1994. – 375 с.
- Гайворонская Г. С. Введение цифрового коммутационного оборудования на местной телефонной сети // Зв'язок. – 2000. – Ч. 1, №4. – С. 31-33.
- Гайворонская Г. С. Некоторые аспекты перехода от аналоговых к цифровым сетям связи // Модели информационных сетей. – М.: Наука, 1984. – С. 145-155.
- Гайворонская Г. С. Проблема синтеза развивающихся информационных сетей // Вісник ДУИКТ. – 2005. – №3. – С. 14-21.
- Гайворонская Г. С. Проблемы выбора ЦСК для создания плоской двухуровневой сети связи Украины // Телеком (телекоммуникации и сети). – 2000. – №7-8. – С. 25-26.
- Гайворонская Г. С. Особенности сетевого планирования с учетом взаимодействия сетей // Вісник УБЕНТЗ. – 2005. – №1. – С. 38-50.
- Гайворонская Г. С. Пути перехода от аналоговых к цифровым сетям связи // Электросвязь. – 1984. – №10. – С. 49-51.
- Гайворонская Г. С. Экономические аспекты оптимизации эволюционной политики развития местных сетей связи:// Сб. междунар. акад. информатизации при ООН Acta Academia 2001. – Chisinau: Evrica, 2001. – С. 93-98.
- Панфилов И. П., Гайворонская Г. С. Оптимизация размещения цифровых систем коммутации на местной телефонной сети // Тр. УГАС "Информатика и связь". – Одесса, 2001. – С. 35-39.
- Гайворонская Г. С. Оптимизация выбора сценария эволюции местной телефонной сети // Зв'язок. – 2002. – №1. – С. 51-53.
- Гайворонская Г. С. Оценка влияния некоторых факторов на процесс развития телекоммуникационных сетей // Холодильна техніка і технологія. – 2006. – №2 (100). – С. 95-100.
- Гайворонская Г. С. Один из подходов к синтезу топологической структуры телекоммуникационной сети // Вісник ДУИКТ. – 2006. – Т. 4, №2. – С. 119-122.
- Гайворонская Г. С. Модель синтеза оптимальной структуры телекоммуникационной сети // Захист інформації. – 2006. – №4. – С. 78-84.
- Гайворонская Г. С., Сомсиков А. С. Исследование некоторых аспектов модернизации телекоммуникационной сети // Сб. междунар. акад. информатизации при ООН Acta Academia 2002. – Chisinau: Evrica, 2002. – С. 155-164.
- Эддоус М., Стенсфилд Р. Методы принятия решений / Пер. с англ. Под ред. И. И. Елисеевой. – М.: ЮНИТИ, Аудит, 1997. – 590 с.
- Пратт Ш. Оценка бизнеса. – М.: Квинто-Консалтинг, 2005. – 392 с.
- Гайворонская Г. С. Метод представления пространственно-временной структуры модернизируемой телекоммуникационной сети // Зв'язок. – 2006. – №8 (68). – С. 57-60.
- Гайворонская Г. С., Сомсиков Д. А. Исследование модели требований на развитие информационной сети // Холодильна техніка і технологія. – 2006. – №3 (101). – С. 99-104.
-

Информация об авторах



Галина Гайворонская – Институт информационных технологий Одесской государственной Академии холода, д.т.н., профессор, заведует кафедрой информационно-коммуникационных технологий; ул. Дворянская, 1/3, Одесса-26, 65026, Украина; тел. (048)-720-91-48, e-mail: gayvoronska@osar.odessa.ua
 Главные области научных исследований: оптимизация переходных периодов при эволюции телекоммуникационных сетей. Потоки вызовов, нагрузка и межузловое тяготение в сетях. Проблемы создания перспективных сетей доступа Проблема построения полностью оптических систем коммутации

ОСОБЕННОСТИ ПРИМЕНЕНИЯ ОПТИЧЕСКИХ КОММУТАТОРОВ В СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ СЕТЯХ

Гайворонская Галина Сергеевна, Рябцов Александр Васильевич

Аннотация: Настоящая статья посвящена вопросам применения коммутаторов оптических сигналов в информационных сетях. Рассмотрены различные методы и средства отклонения и модуляции оптического излучения, на основе которых возможно построение полностью оптических коммутаторов. Проведен краткий анализ технических аспектов построения данных устройств.

Ключевые слова: Оптические коммутаторы, ООО и ОЕО коммутаторы, архитектура коммутаторов, технологии коммутации, параметры коммутаторов.

Ключевые слова классификации ACM: [H.4.3 Communications Applications](#)

Введение

Известно, что главным достоинством оптоволоконных сетей является их практически неограниченная пропускная способность. Практическая ценность этого свойства заключается в возможности многократного увеличения скорости передачи информации по оптоволоконным каналам связи в глобальном масштабе. И это делает исследования в области оптических сетей весьма актуальными и перспективными.

Помимо важной задачи совершенствования параметров и конструкций магистральных волоконно-оптических кабелей не менее остро стоит вопрос создания надежных и доступных по цене коммутаторов оптических сигналов, без которых невозможно построение разветвленных оптических сетей.

Коммутаторы – одни из наиболее важных узлов систем передачи информации, построенных на основе иерархических структур. Без них практически невозможно осуществлять автоматическое управление перемещением потоков данных по разветвленной сети.

Коммутатором в обобщенном смысле обычно называют устройство, позволяющее передать мгновенно и с минимальными искажениями исходный сигнал, поступивший на один из нескольких входов, на один из нескольких выходов в соответствии с заданным алгоритмом переключения. Соответственно под оптическими коммутаторами подразумеваются коммутаторы, осуществляющие коммутацию сигналов, представленных в виде оптического излучения.

Огромное разнообразие применяемых в технике устройств, осуществляющих функцию коммутации оптических сигналов, определяется частными особенностями их применения в конкретном типе сети.

ООО и ОЕО коммутаторы

В настоящий момент в стадии разработки находятся два основных типа коммутаторов оптических сигналов: оптические коммутаторы (all-optical switches), которые называют также прозрачными, световыми (photonic) или ООО (optical-optical-optical) коммутаторами, и оптико-электрооптические коммутаторы, называемые также непрозрачными или ОЕО (optical-electrical-optical) коммутаторами. Полностью оптические ООО-коммутаторы принимают оптический сигнал на входе, осуществляют его коммутацию, и направляют через оптический выход далее без преобразования в электрический сигнал. Оптико-электрооптические ОЕО-коммутаторы принимают оптические сигналы на входе, преобразуют их в электрические сигналы, осуществляют процесс коммутации при помощи электронных компонентов, при этом могут выполнять определенные функции, такие как, например, поиск адресов в базе данных маршрутизации, после чего вновь преобразуют электрический сигнал в оптические сигналы на выходе.

Хотя оба типа оптических коммутаторов преследуют единую цель - увеличение эффективности сетей с помощью гибкой переадресации трафика, разница в их технологических подходах существенна. Каждый из применяемых методов имеет свои достоинства и недостатки, что порождает не только конкуренцию технических решений, но и их технологическую несовместимость. То есть их одновременное внедрение может привести к созданию двух типов оптических сетей, работающих на различных принципах. Именно этот фактор, по нашему мнению, тормозит внедрение новых типов оптических коммутаторов, поскольку ведущие производители сетевого оборудования и сетевые провайдеры до сих пор не пришли к единому мнению, какой из этих двух видов коммутаторов им лучше избрать. Различия в физических принципах действия и в методах управления разными типами коммутаторов, отличия их интерфейсов и отсутствие общепринятых стандартов, и при этом необходимость разработки специализированного программного обеспечения, усложняют выбор. Поэтому в настоящее время мало кто берется предсказать, возможно ли общее решение для всех коммутаторов, применяемых в настоящее время в разнообразных оптических сетях.

В то же время, очевидно, что осуществление коммутации информационных пакетов на фотонном уровне, без промежуточного превращения их в электрические сигналы, позволяет избежать высоких расходов, связанных с преобразованием сигналов, а самое главное, позволяет на несколько порядков увеличить скорость передачи информации.

К сожалению, на сегодняшний день эта идея пока не вышла за пределы исследовательских лабораторий и до ее практической реализации еще весьма далеко. Сегодня еще невозможно обойтись без совместного присутствия в сетях и фотонов, и электронов. Оптико-электрооптические коммутаторы могут извлекать служебную сетевую информацию, например, адресную, и восстанавливать затухающий сигнал. На сегодняшний день ни одна из этих функций не может быть выполнена исключительно оптическими средствами.

Электрооптические методы коммутации применяются, например, в упорядочивающих (grooming) коммутаторах, которые просматривают входящий STM-л-поток, идентифицируют пункты назначения мультиплексированных каналов, а затем реорганизуют каналы таким образом, чтобы сделать передачу максимально эффективной. При этом полностью оптические мультиплексоры лишь объединяют информационные кадры в единый поток и направляют их в выходной порт вне зависимости от того, куда они пересылаются.

Принято считать, что в оптико-электрооптических коммутаторах невозможно добиться высокой плотности портов, в большинстве случаев этот показатель ограничивался 256 портами. Но, благодаря развитию полупроводниковых технологий, сегодня для ОЕО коммутаторов уже достижима плотность, достигающая 1024 портов. Однако, ООО коммутаторы обладают более высокими значениями этого показателя, потенциально достигающего десятков тысяч портов, при соответствующем развитии нанотехнологий. В случае ООО из одного порта в другой перемещается, как правило, вся полоса пропускания волоконно-оптического кабеля. Коммутаторы ОЕО обычно допускают возможность дифференцированного перемещения отдельных волновых каналов. Разработки в области ОЕО-коммутаторов сосредоточены, в основном, на совершенствовании технологии изготовления основных компонентов и их микроминиатюризации. В то же время область исследований полностью оптических коммутаторов второго типа чрезвычайно обширна и открыта для принципиально новых разработок.

И все же, поскольку непрозрачные коммутаторы не в состоянии в полной мере обеспечить те характеристики масштабируемости и независимости протоколов, которые свойственны прозрачным коммутаторам, сосуществование обеих технологий, по всей видимости, будет сохраняться. В то время как непрозрачные коммутаторы применяются для организации доступа к упорядоченному трафику, прозрачные коммутаторы используются в центральной части сети либо для агрегации данных с целью повышения скорости передачи через порт, либо для обеспечения ремаршрутизации трафика в случае отказа одного из сегментов сети без преобразования сигналов в электрическую форму. По мере того как усиливаются возможности оконечных коммутаторов за счет DWDM и интеллектуальной маршрутизации,

эти коммутаторы, обладающие улучшенными возможностями упорядочения трафика, могут обрабатывать все большее число каналов. В результате этого создаются благоприятные условия для расширения масштабов применения методов оптической коммутации в сети.

Технологии создания оптических коммутаторов

За сравнительно короткую историю развития оптических сетей было разработано огромное число различных типов коммутаторов оптических сигналов, основанных на самых различных физических принципах. Главной задачей таких устройств является обеспечение максимально высокого быстродействия, достаточного для разрабатываемых терабитных оптических линий связи, обеспечение минимально возможной удельной стоимости, а также снижение потребляемой мощности по сравнению с существующими электронными аналогами.

Оптические коммутаторы могут быть построены на основе целого ряда известных устройств, таких как электромеханические модуляторы и дефлекторы луча, дифракционные решетки, электрооптические элементы и т.д.

За последнее время значительное число публикаций было посвящено разработке или совершенствованию различных конструкций оптических коммутаторов с электростатическим управлением. В основном это мембранные или лепестковые отражатели, представляющие собой подвижное зеркало на гибкой подвеске, или выполненные в виде гибкого элемента из эластичного непроводящего материала, например, в виде полимерной пленки с высококачественным зеркальным покрытием. Такие отражающие свет поверхности приводятся в движение электростатическими силами, и, обладая малой инерционностью, обеспечивают нужные показатели быстродействия при отклонении световых потоков в оптическом коммутаторе. Однако чувствительность этих конструкций к внешним полям не позволяет им получить широкого распространения в современной технике связи.

Незначительную долю модуляторов светового потока, которые могли бы быть использованы для коммутации оптических каналов в информационных сетях, составляют магнитооптические модуляторы, действие которых основано на эффекте Фарадея. Этот эффект заключается в изменении направления поляризации оптического излучения, проходящего через вещество при наведении в нем магнитного поля. Магнитооптические модуляторы характеризуются большой потребляемой мощностью и малым быстродействием порядка 10-5с, они также подвержены саморазогреву и недостаточно нетехнологичны, поэтому они практически не используются в современной оптической технике.

В настоящее время основную массу устройств, применяемых для модуляции светового потока в выпускаемых промышленностью оптических системах, составляют: управляемые зеркальные и оптомеханические отражатели, термооптические, пузырьковые, жидкокристаллические, электрооптические преобразователи и акустооптические преобразователи, электроголографические коммутируемые решетки Брэгга.

За недавнее время было разработано большое количество микромеханических оптических модуляторов различных типов. Наиболее известны оптические модуляторы, построенные на микрозеркальной технологии с пьезоэлектрической отклоняющей системой (АМА), в которой на двух одинаковых пьезоэлектрических столбиках крепится миниатюрное зеркало. При приложении к столбикам напряжения противоположной полярности зеркало отклоняется, так как один столбик сжимается, а другой расширяется. Угол отклонения зеркала составляет 0,25о, при приложенном напряжении порядка 30 В. Другой примером может служить бистабильный микрозеркальный оптический модулятор (DMD), в виде матрицы ячеек, каждая из которых состоит из микрозеркала, укрепленного на поворотной системе сложной конструкции, и управляющей его отклонением ячейки памяти. Ячейка памяти находится под микромеханической системой и содержит 8 МДП транзисторов. Размер элементарной ячейки DMD матрицы - 17 мкм, угол отклонения зеркала 10о в каждую сторону, время переключения 15 мкс.

Такие зеркальные отражатели, иначе называемые микроэлектромеханическими системами (microelectromechanical systems - MEMS), могут представлять собой сотни и даже тысячи миниатюрных

зеркал, размещенных на одной подложке, которые могут поворачиваться на определенный угол в течение долей миллисекунд. Сфокусированные световые лучи, попадая в эти зеркала, отражаются от них и перенаправляются в соответствующий выходной порт. Коммутаторы на основе MEMS разрабатываются в многоплоскостных вариантах исполнения. Так, например, существуют одно-, двух- и трехплоскостные конструкции, в зависимости от количества координат, в которых отклоняется световой луч.

Эти микроминиатюрные зеркала изготовлены с высочайшим оптическим качеством, что включает в себя, прежде всего, отражательную способность близкую к 100%, малую дисперсию и абберацию. Зеркала перемещаются или вращаются с помощью актуаторов - приводов, основанных на различных физических принципах. К таким актуаторам можно отнести электростатические, электромагнитные и магнитоэлектрические, пьезоэлектрические и магнитоэлектрические приводы. Основным критерий при выборе типа актуатора – максимальное быстродействие при минимальной потребляемой мощности.

Особого внимания на наш взгляд заслуживают пьезоэлектрические актуаторы. К их достоинствам неизменно относят твердость и монолитность, обеспечивающую высокую надежность и долговечность, а также очень высокое входное сопротивление, что позволяет свести практически к нулю входные токи и, соответственно, потребляемую коммутатором мощность.

Основным недостатком таких коммутаторов на основе управляемых зеркальных отражателей MEMS является их низкая механическая прочность. Единственным путем преодоления этого недостатка на сегодняшний день является резервирование путем введения в конструкцию избыточного количества запасных каналов. Еще одним недостатком MEMS является потеря мощности сигнала при его многократных отражениях, достигающая нескольких децибел. Кроме того, пока не удается обеспечить достаточно высокую плотность заполнения, что в принципе, определяется лишь недостаточной проработанностью технологии изготовления. Существуют коммутаторы с конфигурацией 64x64 портов, однако это достигается за счет ухудшения общей надежности. Все еще находится в стадии разработки и вопрос обеспечения точности и стабильности угла поворота зеркала. Использование зеркал, а не электроники, позволяет перенаправлять световые потоки вне зависимости от объема передаваемой информации, что является значительным преимуществом. С другой стороны, такие коммутаторы не в состоянии разбивать потоки данных, т.е. коммутируют весь полный оптический канал данных.

Известны также оптомеханические коммутаторы, в которых для коммутации сигналов от входного порта в выходной наиболее часто используется подвижный отрезок оптического волокна, перемещаемый с помощью механического ползунка. Несмотря на очевидную простоту, такие устройства не нашли широкого применения, так как передвижение коммутирующих элементов должно осуществляться с высочайшей точностью. Даже незначительные перекосы или отклонения волокна от нужного положения приводят к безвозвратной потере информации. Однако этот метод коммутации, тем не менее, находит применение, особенно для уже упоминавшихся целей резервирования, поскольку его проще всего реализовать, и он наиболее дешев. Этот метод используется в основном тогда, когда требуется осуществить обход сбойных компонентов сети и где не требуется частое переключение. Также этот тип оптических коммутаторов часто применяется в некоторых видах испытательного оборудования.

Одной из оригинальных технологий, применяемых при построении оптических коммутаторов является пузырьковая, суть которой заключается в использовании двух наборов кремниевых подложек с продольными полосками. Нижняя подложка имеет слой, состоящий из вытравленных в кремнии продольных микроскопических канавок, пересекающихся друг с другом. Эти канавки заполнены специальной жидкостью, имеющей тот же показатель преломления, что и кремний. Они играют роль волноводов, по которым может распространяться входящий световой поток. На верхней подложке располагаются электроды, протекание токов по которым вызывает нагрев жидкости и образование пузырьков газа в месте пересечения канавок.

В обычных условиях свет беспрепятственно проходит через жидкость. Но когда в одном из пересечений образуется пузырек, входной луч свет, несущий информационный сигнал, отражается от него, изменяет свое направление и направляется в другой канал. В пузырьковых коммутаторах отсутствуют подвижные

части, что обеспечивает значительно более высокую надежность, чем в случае MEMS или оптомеханических коммутаторов. Однако низкое быстродействие, низкая стабильность и потери в тракте передачи сигнала все еще не позволяют таким коммутаторам получить широкое распространение.

Несомненные преимущества высокой надежности из-за отсутствия подвижных частей и пренебрежимо малая потребляемая мощность привели к использованию в оптических коммутационных технологиях жидких кристаллов. Электрическое напряжение, приложенное к жидкому кристаллу, заставляет молекулы изменить свою ориентацию, в результате чего изменяется показатель преломления. В известных в настоящее время оптических жидкокристаллических коммутаторах для переадресации света из одного порта в другой наиболее часто используется явление поляризации, так как известно, что до 50% интенсивности теряется, если проходящий через такой коммутатор свет не был предварительно поляризован. В современных жидкокристаллических коммутаторах входной световой поток попадает в линзу, которая разделяет поляризованный сигнал на два пучка, обладающие противоположной поляризацией. Оба пучка отражаются от жидкокристаллических слоев и собираются на другой линзу, где они рекомбинируют. На каждый из жидкокристаллических элементов нанесены электроды, меняя величину напряжения на которых, можно изменять поляризацию оптических сигналов, направляя их в нужный выходной порт.

Существенным недостатком таких модуляторов является низкий коэффициент передачи светового потока в канале оптического коммутатора.

В оптической технике применяются еще два способа применения жидких кристаллов для отклонения светового луча. В первой технологии используется дисперсионный жидкий кристалл (PDLC), представляющий собой полимерную матрицу с пузырьками, заполненными молекулами ЖК. Ориентация молекул в разных пузырьках различна, и в отсутствие электрического поля свет, проходящий через матрицу, рассеивается. При приложении к матрице напряжения молекулы жидкого кристалла во всех пузырьках разворачиваются вдоль электрического поля, и свет проходит с минимальным рассеянием. Во второй технологии изготовления оптических коммутаторов на жидких кристаллах мономолекулярный слой жидкого кристалла возбуждается таким образом, чтобы в участках, соответствующих светлым элементам изображения, молекулы образовывали дифракционную решетку, на которой рассеивается входной оптический сигнал.

Однако описанные технологии не получили распространения на практике вследствие явлений температурной нестабильности и фотодеградациии жидких кристаллов, а также взаимного влияния соседних активных областей в кристалле.

В настоящее время жидкокристаллические оптические модуляторы проходящего или отраженного света являются самыми распространенными устройствами в области формирования видеоизображений. К сожалению, в коммутационных технологиях жидкие кристаллы, несмотря на массу своих достоинств, не нашли широкого применения из-за высокой инерционности ЖК молекул, которая обычно составляет от 1 до 10 мс.

Похожая технология применяется и в электроголографических ячейках на основе решеток Брэгга, где управление появлением и исчезновением решетки осуществляется с помощью электрического поля. Для этих целей используются голографические изображения решетки Брэгга, которые создаются в специальных кристаллах, называемых KLTN (potassium lithium tantalate niobate — танталат-ниобат лития-калия). Эти кристаллы образуют ряды и столбцы, где ряды представляют отдельные оптические волокна, а столбцы — волновые каналы. Каждый кристалл управляется электрическим полем с помощью нанесенных электродов. При отсутствии напряжения свет беспрепятственно проходит через все кристаллы, но когда прикладывается напряжение и активизируется голограмма, входящий световой сигнал отклоняется в определенный выходной порт. В основе второй технологии, называемой электрически коммутируемыми решетками Брэгга (electrically switchable Bragg gratings или ESBGs) лежит тот же принцип. Однако чтобы заставить решетку «появляться», слой жидкого кристалла наносится на полимерный слой, покрывающий оптический световод. При отсутствии напряжения решетка отклоняет

распространяющиеся по световоду волны определенной длины. При подаче электрического напряжения решетка исчезает, и свет проходит по световоду без потерь. В электроголографических коммутаторах также отсутствуют подвижные части, а скорость коммутации в ESBG составляет примерно 100 микросекунд, что делает ESBG значительно более быстрыми, чем MEM или пузырьковые коммутаторы. Световые потери составляют порядка 1 дБ. Однако такие коммутаторы характеризуются высоким энергопотреблением, порядка 50 милливатт на коммутационный элемент.

К настоящему моменту основное внимание разработчиков оптической коммутационной техники сосредоточено на создании и совершенствовании электрооптических и акустооптических модуляторов. В полностью оптических коммуникационных сетях AON для коммутации информационных потоков в основном используется электрооптическая модуляция как наиболее эффективный способ управления световым потоком.

Электрооптические коммутаторы, выполненные на основе ячеек Керра или Поккельса, обычно содержат активные участки с электродами, к которым приложено электрическое поле. Эффект Поккельса проявляется в том, что при воздействии электрического поля на некоторые оптически прозрачные кристаллы происходит изменение показателя преломления материала, и соответственно меняется интенсивность света, прошедшего через кристалл. В оптоволоконной технике широко применяются X-ответвители на основе сплавных ячеек Поккельса. При отсутствии напряжения на электродах ячейки эффективность связи между волноводами составляет 100%. Это значит, что оптические сигналы полностью кроссуются, т.е. входят в один волновод, а выходят из другого. При подаче соответствующего напряжения на электроды эффективность передачи уменьшается до нуля. Модулируя величину напряжения на электродах, можно менять такие параметры излучения, как интенсивность света в канале или его фазу, блокируя отдельные каналы и перенаправляя тем самым информацию в виде оптического излучения из одного порта в другой.

Преимуществом электрооптического эффекта является твердотельность конструкции и высокая скорость коммутации. А недостатками является высокое напряжение на электродах, достигающее тысяч вольт. Особые сложности возникают при построении неблокирующих коммутаторов большой размерности.

На рынке сетевых технологий представлены также термооптические коммутаторы, в которых для изменения характеристик оптического волокна применяется температурное воздействие, позволяющее добиться того же эффекта, что и в электрооптических коммутаторах. К ним относятся интерферометрические и цифровые оптические коммутаторы (digital optical switches — DOS). Простейшими являются коммутаторы со структурой портов 1 x 2. Увеличение температуры в одном плече вызывает увеличение коэффициента затухания и блокирует прохождение света в этом плече. Термооптические коммутаторы имеют меньшие размеры, отличаются большей надежностью, но их эксплуатационные характеристики зависят от длины волны пропускаемого света, вследствие чего требуется применение дополнительных средств контроля, обеспечивающих поддержание температуры отдельных соединений в строго установленных пределах.

В последнее время активизировались разработки в области создания оптических модуляторов, основанных на изменении коэффициента преломления полупроводниковых материалов, и в частности кремния, который является оптически прозрачным в типичном для оптоволоконных коммуникаций диапазоне длин волн (от 1,3 до 1,55 мкм) и обладает высоким коэффициентом преломления.

В настоящее время разработаны микроминиатюрные модуляторы светового потока на основе кольцевых кремниевых световодных структур. Световоды выполнены методом плазменного напыления PECVD (Plasma Enhanced Chemical Vapor Deposition) и изолированы стенками из двуокиси кремния. Входной оптический сигнал, в виде линейно поляризованного когерентного излучения с заданной длиной волны, направляется по линейному световоду, выполненному на подложке из кристаллического кремния. Центральная часть световода имеет меньший коэффициент преломления, чем границы, а его геометрические размеры соответствуют длине волны распространяющегося излучения (1,9 мкм).

Главная деталь модулятора - кольцевой резонатор. В нем может распространяться только пучок света с фиксированной длиной волны, которая определяется исходя из геометрических размеров кольца. Было доказано, что если расположить два волновода достаточно близко друг к другу, то энергия может передаваться между ними так, словно разделяющего их пространства (материала, барьера) не существует. При этом имеется возможность подстраивать частоту резонанса и полосу пропускания кремниевого световода с помощью второго пучка света, называемого «накачивающим». Так как кольцо резонатора располагается в непосредственной близости от линейного волновода, то свет проникает через разделяющее их пространство. При этом излучение с резонансной длиной волны будет "поглощаться". Соединяя посредством резонатора два параллельных волновода, имеется возможность получить "оптический вентиль", пропускающий между ними свет только на частотах, кратных резонансной.

Поскольку на частоте резонанса оптическая пропускная способность кольца составляет более 80%, а для других длин волн её значение снижается ниже 5%, то, изменяя таким образом коэффициент передачи, можно осуществлять модуляцию оптических сигналов с заданной длиной волны. Конфигурация из нескольких волноводов и колец, настроенных на разные частоты резонанса, в совокупности может представлять собой оптический многоканальный коммутатор.

К очевидным недостаткам такого устройства можно отнести высокую зависимость полупроводниковых материалов от температуры (как известно, коэффициент преломления кремния нелинейно уменьшается с ростом температуры), а также быстрое действие, ограниченное временем релаксации свободных носителей в кремнии, равным 450 пс.

Наиболее перспективными в настоящее время с точки зрения быстрого действия представляются акустооптические коммутаторы, основанные на эффектах Джекффри или Брэгга, в которых используются акустические волны для создания в оптически прозрачном кристалле областей с повышенной и пониженной плотностью. Образующиеся под действием ультразвука чередующиеся неоднородности в плотности кристалла образуют своего рода дифракционную решетку, которая может отклонять проходящие световые пучки под различными углами. Изменяя картину акустического поля, можно перемещать оптические сигналы между различными портами. Эта технология обеспечивает интервал времени переключения между портами от 500 нс до 10 мкс при полном отсутствии подвижных частей. Недостатком таких коммутаторов является чувствительность к акустическим и температурным шумам.

Еще одним хорошо зарекомендовавшим себя средством коммутации оптических сигналов является использование усовершенствованных элементов традиционной дискретной оптики, например таких, как дифракционные решетки.

В настоящий момент одновременно развиваются три конкурирующие технологии.

Первая применяется в многоканальных мультиплексорах/демультиплексорах и использует выделение оптических несущих на основе дифракционной решетки AWG (Arrayed Waveguide Grating), образованной массивом волноводов разной длины в объеме оптически прозрачной пластины. Свет распространяется по массиву волноводов с разным фазовым запаздыванием и отражается в зеркале на торце пластины, при этом происходит интерференция входной и отраженных волн. Размеры и форма самой пластины и массива внутренних световодов выбираются так, чтобы интерференционные максимумы располагались в районе выходного порта. При демультиплексировании происходит обратный процесс. Входной оптический поток, образованный совокупностью излучений разных длин волн, подается во входной оптический волновод и распределяется по всем внутренним портам, откуда, отразившись от зеркала, распространяется по выходным световодам.

В основе второй технологии лежит применение изогнутой дифракционной решетки CG (Concave Grating), которая позволяет пространственно разнести входы и выходы такого коммутационного узла. При этом для фокусирования и интерференционного разделения каналов используется не одна, а две волноводных пластины. Процессы, происходящие в таком мультиплексоре, аналогичны вышеописанным.

В третьей технологии применяется миниатюрная дискретная оптика, использующая трехмерное оптическое мультиплексирование – 3DO (3-D Optics WDM) с классической оптической схемой,

включающей плоскую отражательную дифракционную решетку, вогнутое зеркало и массив волокон, размещенных в пазах решетки с фиксированным шагом. Принцип работы устройства в режиме демультиплексора: мультиплексированный поток из входного волокна, расходясь конусом, отражается от зеркала и падает на дифракционную решетку, отражающую под разными углами свет разной длины волны. Эти дифрагированные лучи, отражаясь от зеркала, фокусируются в определенных точках, где расположены приемные порты массива волокон, выделяющих соответствующие несущие.

К сожалению, все эти рассмотренные конструкции обладают хорошо известным недостатком прецизионных оптических устройств – низкой эксплуатационной надежностью, необходимостью тщательной юстировки и высокой стоимостью.

Большинство из существующих в настоящее время оптических сетей, а соответственно и коммутаторов оптических сигналов для них, работают с постоянной длиной волны. Однако в настоящее время существует настоятельная необходимость иметь возможность менять длину волны сигналов в процессе их передачи. Поэтому в ближайшем будущем следует ожидать широкого развертывания систем передачи с преобразованием длин волн. И в этой связи разработка коммутаторов с преобразованием длин волн является весьма актуальной задачей. Возможно, наиболее перспективными в этом качестве могут явиться OADM (optical add-drop multiplexers), которые позволяют извлекать или добавлять в линии DWDM отдельные волновые каналы. В настоящее время они используются, в основном, в качестве упорядочивающих коммутаторов, или агрегируют низкоскоростной трафик для передачи его по каналам OC-3 или OC-12. В центральной части общегородских сетей они выступают в роли оптических кросс-соединителей, осуществляющих подключение множества входящих линий, и могут работать с частью полосы пропускания отдельных волоконно-оптических кабелей, полосами пропускания и целыми волновыми каналами.

Архитектура оптических коммутаторов

Современные оптические коммутаторы в основном строятся на основе двух структур. Для более простой из них характерно использование координатных (матричных) коммутаторов. Линии входных и выходных оптических каналов образуют прямоугольную сетку. В случае ОЕО коммутаторов эта сеть обычно образована электронными ключами. В полностью оптических коммутаторах вместо электронных ключей применяются оптические ответвители (optical couplers). При этом различные порты соединяются между собой путем задания различных положений этих ответвителей.

В таких координатных коммутаторах основные проблемы связаны с большим размером матрицы коммутирующих элементов, т.е. числа электронных или оптических ключей, достигающих в современных коммутаторах десятков или даже сотен тысяч. Для каждого из этих пересечений требуется делитель (splitter) или объединитель (combiner), что увеличивает стоимость системы. В связи с этим координатные коммутаторы используются в основном в качестве конструктивных матриц небольшой емкости при разработке иерархических коммутаторов увеличенной емкости. В случае коммутаторов подобного конструктивного типа, называемого архитектурой Клоза (Clos architecture), имеется три яруса соединений. В такой структуре входные порты подключены к одному набору координатных коммутаторов, а выходные — к другому. И при этом и входной и выходной наборы подключены к третьему.

При использовании транспозиционных коммутаторов архитектура Клоза может быть использована в коммутационной структуре Бенеша (Benes switch fabric), где коммутаторы с конфигурацией 2×2 применяются для осуществления произвольных комбинаций. При этом половина портов любого из промежуточных коммутаторов Клоза расположена с одной стороны коммутатора и играет роль входных портов, а вторая, расположенная с другой стороны, — выходных. Каждый из этих коммутаторов также может быть разбит на две группы, и так далее, пока вся схема не будет образована соединителями 2×2 .

Однако, при том, что архитектура Бенеша достаточно близка к оптимальной с точки зрения использования аппаратных средств, в таких матричных коммутаторах все же целесообразнее использовать

коммутирующие элементы с размерностью, большей чем 2,2 иначе при большом числе каналов возникают проблемы с масштабированием. Коммутаторы общего типа со сходными архитектурами имеют аналогичные недостатки из-за еще большего числа коммутирующих элементов. Кроме того координатным коммутаторам свойственно ослабление сигналов, происходящее при двухстадийном процессе разделения и объединения сигналов, что требует использования энергетически управляемого разделения и объединения (power-dividing and power-combining) сигналов.

Основные функции коммутаторов

Все современные оптические коммутаторы, как ООС, так и ОЕО, выполняют в сети несколько ключевых функций. К ним можно отнести резервирование, собственно процессы коммутации и функции контроля.

Суть оптического резервирования практически не отличается от традиционной задачи автоматической защиты сетевого трафика от любых сбоев в линейных компонентах сети (в данном случае - от обрывов волоконно-оптических кабелей или отказов EDFA-усилителей). Оптическое резервирование обеспечиваются заблаговременным размещением таких коммутаторов в местах критических соединений сети с созданием возможности обходных путей.

При изменении направления потока информации в оптической сети коммутатор должен решать две важные задачи. В коммутаторах оптического мультиплексирования методом добавления/ответвления каналов (optical add-drop multiplexers - OADM) коммутирующие матрицы используются для добавления или извлечения данных, передаваемых по магистральным волновым каналам (wavelength). Краевые (Edge) коммутаторы используются для переключения входящих низкоскоростных линий для передачи данных в магистрали OC-3 или OC-12. Кросс-соединители (photonic crossconnects), называемые также оптическими коммутаторами (photonic switches) или маршрутизаторами по длине волны (wavelength routing), размещаются в местах критических соединений в пределах центральной части сети, коммутируя из одного порта в другой всю полосу пропускания волоконно-оптического кабеля или его отдельные волновые каналы.

Оптические коммутаторы могут быть также использованы для проверки качества отдельных компонентов сетей или для контроля активности трафика путем добавления в трафик некоего тестового сигнала, не нарушающего прохождения основных информационных потоков. Например, в таких испытательных приборах, как оптические рефлектометры, коммутаторы используются для проверки множества волоконно-оптических кабелей на удаленных узлах.

Наиболее распространенными ООС коммутаторами являются оптические кросс-соединители (optical crossconnects — ОКС), которые принимают данные от одного из входных портов и коммутирует их на два или более выходных.

Сейчас в оптических сетях используются два типа кросс-соединителей, или, как их принято называть, пространственных (space-division) коммутаторов: транспозиционные (permutation) и общего типа (generalized). Транспозиционные коммутаторы, или оптические прерыватели, состоят из соединений типа «один вход - один выход» между различными портами. Соединение типа «один вход-множество выходов» для данного типа коммутаторов невозможно. При этом общее число возможных кросс-соединений определяется количеством транспозиций, т.е. суммарным количеством попарных комбинаций возможных выходов и входов.

Второй тип кросс-коннекторов может соединять любой входной порт с множеством выходных или же наоборот. При этом число различных возможных комбинаций соединения портов увеличивается. Их разновидностью являются специальные коммутаторы, называемые линейными делителями-объединителями (linear divider-combiner - LDC), которые позволяют равномернее распределять входную мощность оптического сигнала между выходными портами для обеспечения улучшенных характеристик затухания.

Основными недостатками таких кросс-коннекторов является то, что каждый входной оптический сигнал должен обязательно пройти на выход через какой-либо выходной порт коммутатора во избежание его

отражения обратно ко входу. Если коммутаторы общего типа и LDC в большинстве случаев позволяют это сделать, то в случае транспозиционных коммутаторов это невозможно. При этом сети, содержащие транспозиционные коммутаторы, должны обладать сложной структурой, предотвращающей непреднамеренные оптические соединения соседних каналов, например, замкнутые кольцевые маршруты в случае сетей.

Кроме того в таких коммутаторах необходимы оптические усилители, являющиеся источником шумов, и в случае образования кольцевых маршрутов эти шумы накапливаются и достигают значительной величины, а также интерферируют с другими волновыми каналами.

Основные характеристики и параметры коммутаторов

Ключевыми параметрами, применяемыми при оценке и анализе технических и эксплуатационных характеристик коммутаторов, являются: размерность коммутационной матрицы, масштабируемость, структурируемость (granularity) и скорость коммутации, а также вносимые потери.

Размерность матрицы — В современных городских условиях потенциальное число линий связи, подключенных к одному коммутационному устройству, может достигать тысяч и даже десятков тысяч. Поэтому количество портов, которые способен поддерживать коммутатор, является его важнейшей характеристикой. Некоторые современные оптические коммутаторы достигают конфигурации 4000 x 4000, однако для подавляющего большинства приложений не требуется такой размерности. Поэтому основная масса коммутаторов рассчитана на гораздо меньшее количество портов, например 16x16 или 64 x 64 порта.

Масштабируемость — Многие сетевые провайдеры используют архитектуры или системы коммутации, для работы которых необходимо несколько десятков портов, например схемы Клоза и Бенеша. Если имеющаяся в настоящее время система коммутации построена на базе коммутационных элементов, имеющих малую размерность, например 2 x 2, то при её масштабировании до большего количества портов, вызванном, например, увеличением числа потребителей услуг, коммутационное устройство придется дополнительно дооборудовать сложными управляющими элементами, что может привести к неоправданным расходам.

Структурируемость — При построении оптического коммутационного устройства особую важность приобретают вопросы, связанные с размерами портов и их согласованием. Весьма желательно, чтобы система поддерживала все типы портов (OC-3, OC-12, OC-48, OC-192, OC-768). Число портов имеет значение не только в случае использования оптических кросс-соединителей, но и при маршрутизации волновых каналов, когда ключевой характеристикой является число коммутируемых каналов, а не только число подключенных кабелей.

Скорость коммутации — Сейчас практически невозможно точно ответить, какова должна быть реальная скорость перемещения светового потока из одного порта коммутатора в другой.

Этот параметр в основном ограничивается допустимым уровнем финансовых затрат. В настоящее время считается вполне достаточным, если коммутация осуществляется в течение нескольких миллисекунд. SONET проигрывает в этом отношении, обеспечивая поддержку соединений со скоростью 50 миллисекунд, в то же время, обеспечивая преимущество по сравнению с протоколами более высокого уровня.

Известные в настоящее время полностью оптические коммутаторы уверенно работают при скоростях 40-160 Gb/s. В то же время перспектива создания терабитных сетей требует переключения отдельных пакетов оптическими средствами с пикосекундными периодами коммутации.

Потери — В оптических коммутаторах присутствуют два основных вида потерь. Первые вносятся в месте сопряжения волоконно-оптических кабелей с коммутатором, а вторые определяются свойствами самого коммутационного элемента или коммутационной матрицы. Несмотря на устойчивую тенденцию к минимизации, потери в современных оптических коммутаторах могут составлять от 1-2 дБ до 4-5 дБ.

Помимо вышеупомянутых немаловажными показателями эффективности коммутатора при его выборе могут также служить переходные и перекрестные искажения, вносимые в исходный оптический сигнал паразитными явлениями рефракции, дисперсии и интерференции. Немаловажными критериями, определяющими выбор типа коммутатора, являются также массо-габаритные показатели, удельная стоимость, приходящаяся на один коммутируемый канал, а также надежность.

Одной из важнейших характеристик коммутаторов является характеристика блокирования. Этим термином обычно обозначается невозможность, по каким либо причинам, обеспечить одновременное соединение между произвольно взятыми входным и выходным портами.

В некоторых случаях, в зависимости от структуры коммутатора или быстродействия коммутирующего элемента, не удастся организовать некоторые межсоединения, и они оказываются заблокированными.

В настоящее время известны несколько видов неблокирующих коммутаторов, и каждый из них обладает определенными недостатками. Например, в неблокирующих коммутаторах с изменяемой структурой (rearrangeably nonblocking switches) отдельные элементы коммутатора, например, отрезки световода, механически перемещаются так, чтобы создать новое соединение. Однако не всегда возможно осуществления нового соединения без прерывания одного или нескольких уже установленных соединений.

Матричные коммутаторы вида $N \times M$ также часто относят к числу перестраиваемых неблокирующих коммутаторов. Но при этом задание непересекающегося пути следования оптического сигнала по коммутационной матрице не всегда просто осуществить. Вследствие этого, несмотря на сравнительную простоту аппаратной части, сложность управления остается довольно высокой. Неблокирующие коммутаторы с архитектурой Клоза позволяют избежать временного отключения существующих соединений лишь в том случае, когда для установления нового соединения использован правильный алгоритм переключения. Но для осуществления сложных алгоритмов маршрутизации требуется дополнительное оборудование и соответствующее программное обеспечение. В коммутаторах Бенеша для установления новых соединений может быть выбран любой свободный маршрут. Но при этом резко возрастает стоимость оборудования в связи с увеличением общего количества коммутационных элементов.

Проблема совместимости оптических коммутаторов

В настоящее время существуют два глобальных подхода к построению систем оптической коммутации. Одним из подходов к задаче внедрения оптических коммутаторов является полная замена существующей структуры уже работающих электронных коммутаторов их оптическими аналогами. При этом замене подлежат все элементы существующей сети, а также применяемые протоколы и программное обеспечение.

Второй подход заключается в постепенной замене существующих электронных коммутаторов на их оптические аналоги с увеличением быстродействия и пропускной способности, при сохранении прежней структуры сети. При этом сохраняется и существующая иерархия применяемых коммутаторов, которая включает в себя настольные коммутаторы, предназначенные для работы с небольшим числом пользователей, магистральные коммутаторы и коммутаторы для рабочих групп.

Настольные оптические коммутаторы должны послужить заменой используемых концентраторов 10Base-T. Для этого они будут иметь до 24 портов, каждый из которых поддерживает персональный (private) канал (например, для рабочей станции). Дополнительно такой коммутатор должен иметь один или несколько портов, аналогичных нынешним 100Base-T или FDDI, для подключения к магистрали (backbone) или серверу. Настольные коммутаторы должны быть также как и сейчас просты в установке и обслуживании.

Магистральные коммутаторы применяются в основном для соединения сетей или сегментов, поддерживающих множественную адресацию для своих портов. Такие коммутаторы будут использоваться для соединения настольных оптических коммутаторов, аналогичных концентраторам 10Base-T, групповых

коммутаторов и серверов. Магистральные коммутаторы должны одновременно передавать трафик между несколькими сегментами с полным использованием полосы пропускания среды. Для магистральных коммутаторов будет характерно модульное устройство и способность поддерживать до нескольких тысяч MAC-адресов на каждый порт. Установка таких коммутаторов более сложна по сравнению с настольными коммутаторами, за счет необходимости настройки функций маршрутизации. Так же как и в современных электронных устройствах такого типа, обязательными будут являться резервные источники питания, горячая замена модулей, поддержка всех видов протоколов для магистральных коммутаторов, обеспечение всех известных возможностей и технологий коммутации, как например виртуальные сети.

Коммутаторы для рабочих групп, которые сейчас используются главным образом для соединения изолированных настольных коммутаторов или концентраторов 10Base-T с остальными частями сети. Эти устройства объединяют в себе свойства как настольных, так и магистральных коммутаторов. Подобно магистральным, коммутаторы рабочих групп должны будут поддерживать множественную адресацию (до нескольких тысяч MAC-адресов на коммутатор) и позволять использование в качестве маршрутизаторов. Как и настольные коммутаторы они могут служить для подключения к портам отдельных узлов. Это позволит пользователям перейти от работы с разделяемой средой к персональным (private) каналам за счет одновременной передачи по оптическому кабелю нескольких волновых каналов

Заключение

Приведенный в данном докладе обзор существующих в настоящее время оптических коммутаторов и тенденций их развития позволяет сделать однозначный вывод, что несмотря на широкое разнообразие физических принципов, на которых построены известные на сегодня средства коммутации световых потоков, ни одно из них пока не способно полностью удовлетворить всех требований, предъявляемых к узлам современных информационных сетей, а именно: высочайшей надежности, достаточному быстродействию, высокой технологичности и соответственно, целесообразной стоимости. Данный обзор не претендует на всеобъемлющую полноту, но может быть полезен для анализа современного состояния проблемы создания полностью оптических информационных сетей и выбора направлений исследований.

Библиография

Авторы не приводят список литературы к докладу ввиду его значительного объема.

Информация об авторах



Гайворонская Галина Сергеевна – д.т.н., профессор, заведующая кафедрой ИКТ Одесской Государственной Академии Холода, Одесса, Украина.

e-mail: gayvoronska@osar.odessa.ua

Главные области научного исследования: общие теоретические исследования в информационных технологиях, информационные и телекоммуникационные системы.



Рябцов Александр Васильевич – к.т.н., доцент кафедры ИКТ Одесской Государственной Академии Холода, Одесса, Украина.

e-mail: ryabtsov@usa.com

Главные области научного исследования: информационные системы, коммутационная техника, оптические технологии, полностью оптические коммутаторы.

ОЦЕНКА ЧУВСТВИТЕЛЬНОСТИ ХАРАКТЕРИСТИК СЕТЕЙ ДОСТУПА К ВАРИАЦИЯМ ПРОГНОЗИРУЕМЫХ ПАРАМЕТРОВ

Светлана Сахарова

Аннотация: Рассмотрена связь между прогнозируемыми входными параметрами и характеристиками сети доступа. Предложен метод оценки чувствительности сети к вариациям прогнозируемых параметров с помощью моделей прямоугольной и секторной структуры сети доступа. Разработаны алгоритм и программный продукт реализации метода.

Ключевые слова: сеть доступа, узлы доступа, прямоугольная и радиальная модели сети доступа, прогнозируемые параметры сети доступа.

Ключевые слова классификации АСМ: С.2. Computer-communication networks, H. Information Systems - H.1 Models and Principles, K. Computing Milieux - K.6 Management of computing and information system

Введение

При модернизации существующих или создании новых современных телекоммуникационных сетей (ТС), важнейшим этапом является этап проектирования. Изучение того, как вариации исходных параметров влияют на результаты сетевого планирования, является одной из основных задач проектировщика. Важность этой задачи определяется тем, что при проектировании ТС в качестве исходных параметров задаются некоторые прогнозируемые данные, характеризующиеся определенной неточностью. К прогнозируемым параметрам относятся распределение пользователей инфокоммуникационных услуг (ИКУ) по обслуживаемой территории, и удельная нагрузка, создаваемая различными группами пользователей. Оптимальная сетевая структура, а соответственно и ее стоимость, будут меняться в зависимости от вариации прогнозируемых параметров. Причем даже незначительное изменение величины исходных данных может существенно повлиять на выходные характеристики. Стоимостные параметры оборудования, которые также являются входными данными, меняются согласно технологическим новшествам и их возможное изменение также следует учесть при проектировании. Поэтому исследование вопроса, каким образом вариации исходных данных влияют на конфигурацию сети, и ее стоимость, является весьма важным.

Исследование, результаты которого приведены в данной работе, выполнены для перспективных сетей доступа (СД). Проектирование СД в настоящее время приобрело особую актуальность, поскольку именно СД является тем сегментом ТС, который тормозит внедрение широкополосных высококачественных услуг, за счет которых оператор может иметь значительный доход. Кроме того, модернизация абонентских линий и создания на их базе перспективных СД является третьим этапом преобразования ТС, первыми двумя этапами которого является замена аналоговых систем передачи и аналоговых узлов коммутации на цифровые. Первый и второй этапы модернизации ТС осуществляются возрастающими темпами, а сейчас пришло время третьего заключительного этапа, который непосредственно связан с темой данной статьи. Структура перспективных СД определена в ряде рекомендаций МСЭ, в частности в рэк. G 902, и подробно рассмотрена в [1].

Постановка задачи

Представленная работа, направленная на оценку чувствительности характеристик СД на вариации прогнозируемых входных параметров и анализ влияния этих вариаций на технические и стоимостные

характеристики сети. Целью работы является повышение точности проектирования сетей доступа, выявление тех параметров, которые существенно влияют на результаты проектирования СД. В работе решается ряд задач, среди которых:

- определение набора параметров, которые влияют на структуру и стоимость сети доступа, представление этих параметров в форму, удобную для моделирования сети;
- на основе выбранных параметров построение математической модели СД;
- разработка методики исследования влияния изменений величины прогнозируемых параметров на характеристики моделируемой СД;
- разработка алгоритма и программного продукта, реализующего модель создания СД;
- путем моделирования процесса создания СД оценка чувствительности характеристик СД на вариации прогнозируемых параметров;
- анализ полученных результатов.

Выбор и группировка параметров СД

При проектировании СД первоначальный этап отводится предпроектным изысканиям, цель которых максимально полный сбор исходных данных для проектирования сетей. На данном этапе рассматриваются всевозможные параметры сетей доступа, составляется характеристика и описание каждого из них, выявляется корреляция и анализируется взаимосвязь между ними. Проводится выбор тех параметров, которые являются наиболее значимыми при моделировании сетей доступа, значения которых существенно влияют на структуру СД. Также выделяются параметры, не являющиеся важными, и в рассмотрении которых при проектировании СД нет особой необходимости. Учитывая то, что при рассмотрении как проводных так и беспроводных систем доступа к ИКУ выделено достаточно большое количество параметров, для удобства выбранные параметры были разделены на группы.

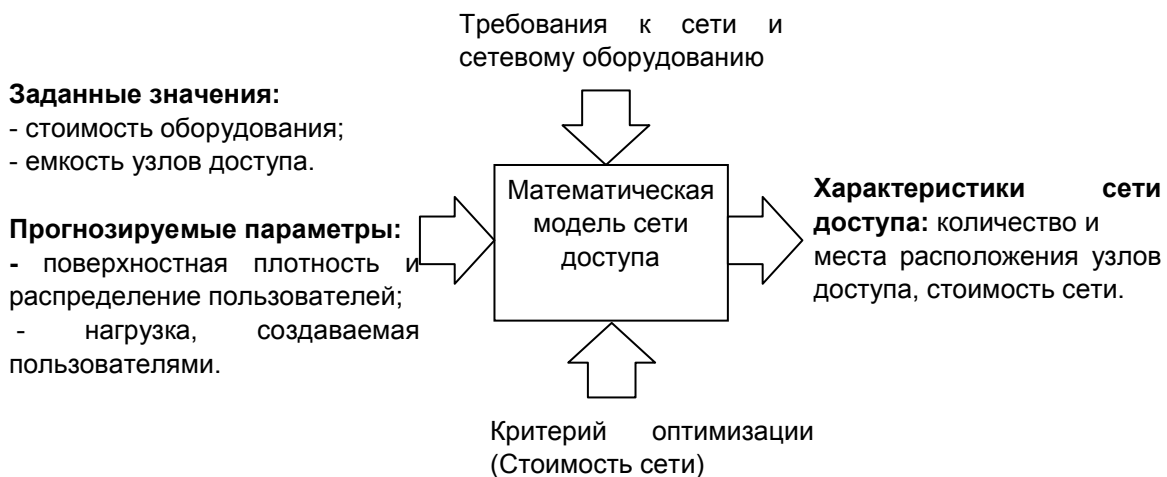


Рис.1. Выбор и группировка параметров СД

В работе предложен один из способов разделения параметров на группы.

1. Параметры, не являющиеся параметрами сети доступа, но влияющие на процесс моделирования. К ним относятся, например, финансовые возможности пользователей, тип местности, ее географическое положение и т.д.
2. Параметры сети доступа, являющиеся входными данными для моделирования. Среди них выделяются:

- требования пользователей, на которых ориентирована создаваемая СД, такие как скорость передачи информации, дальность связи, коэффициент ошибок, время задержки, вариация задержки и т.п.;
- параметры, основанные на требованиях пользователей, но не являющиеся важными для них, такие как метод модуляции в канале, применяемая технология доступа, метод разделения канала приема и передачи и т.п.

В зависимости от набора вышеперечисленных требований и параметров формируются такие параметры, как стоимость услуг, качество связи.

Выбранные параметры требуется представить в форме, удобной для моделирования. Весь набор параметров делится на две категории: качественные и количественные параметры. Для количественных определяются диапазоны принимаемых значений. Для качественных так же определена градация возможных значений.

Для распределения параметров по их важности в процессе проектирования выделяются признаки, по которым можно составить классификацию параметров. Как пример, можно провести классификацию параметров для СД по классу обслуживаемой местности, т.к. при проектировании сети необходимо учесть особенности той местности, для которой она строится. Учитываются такие параметры, как: географические особенности местности, плотность населения, поверхностная плотность пользователей, их финансовые возможности, тип местности.

В работе, при создании модели СД, основное внимание уделено параметрам, значения которых не являются точно заданными, а прогнозируются в процессе моделирования. Ошибки прогноза, в таком случае, могут существенно повлиять на характеристики проектируемой СД. В процессе моделирования оценивается влияние отклонения возможного значения параметра от предполагаемого значения.

В качестве прогнозируемых параметров в работе выбраны: поверхностная плотность пользователей на территории, обслуживаемой СД, распределение пользователей по территории, нагрузка, создаваемая различными группами пользователей.

Модель исследования

Моделирование процесса создания СД заключается в получении характеристик СД на основании заданных прогнозируемых параметров. Для этого необходимо разработать математическую модель СД. При проектировании СД возможны два подхода: модернизация существующей абонентской сети (АС) или синтез новой сети. В любом случае возникает необходимость в исследовании структуры и методов расчета АС. Математическая модель СД, предложенная в работе и подробно описанная автором в [2-4], построена с учетом особенностей сетей доступа, при этом были учтены и использованы принципы расчета АС. В связи с существенными различиями структур АС и СД методы, предложенные для расчета АС, подверглись переработке и коррекции, и ориентированы на определение мест расположения узлов доступа (УД). Формулы, предложенные в работе, адаптированы к структуре СД, отличающейся наличием УД, появлением двух сегментов линий доступа (ЛД): транспортного и локального, и изменением функций ЛД по сравнению с функциями абонентских линий телефонных сетей. Эти различия подробно проанализированы в [5].

Модель предусматривает различные варианты построения СД в зависимости от формы территории, которая будет ей обслуживаться, формы участков, обслуживаемых УД и способов прокладки ЛД. В статье рассматриваются прямоугольная и радиальная модели СД (рис. 2). Эти модели существенно отличаются математическими методами расчета характеристик сети.

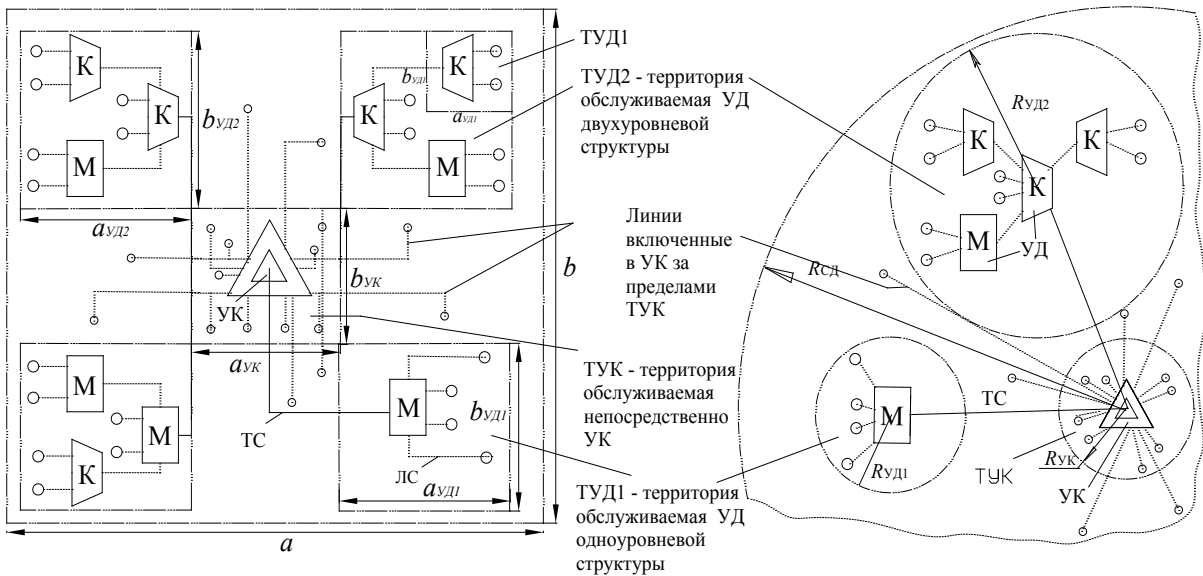


Рис. 2 Прямоугольная и радиальная модели сети доступа

Модель прямоугольной территории СД, учитывающая способы застройки принятые в крупных городах, характеризуется ортогональной прокладкой линий, однородной плотностью размещения пользователей и прямоугольными территориями, обслуживаемыми одним УД (ТУД). Модель секторной территории СД, разработанная при анализе АС в малых городах, характеризуется радиальной прокладкой линий, разнородной плотностью распределения пользователей и трапецеидальными формами ТУД. Предполагается, что узлы предоставления услуг (УПУ) располагаются за пределами территории СД, а СД подключается ко всем УПУ посредством узла коммутации (УК). Территория, в пределах которой пользователи подключаются непосредственно к УК (ТУК), имеет форму, соответствующую модели.

Одной из основных задач при построении модели СД является определение мест расположения УД. Оптимальное местоположение УД определяется исходя из экономических соображений. Узлы доступа, используемые в этой модели, реализованы в виде концентраторов (К) или мультиплексоров (М). Линии, на которых используются УД могут иметь разноуровневую структуру, при этом возможны различные комбинации УД. Территории, обслуживаемые УД одноуровневой структуры обозначаются как ТУД1, двухуровневой – соответственно ТУД2. Подключение УД к УК реализовано с помощью транспортного сегмента ЛД, обозначенного на рисунке ТС, соответственно локальный сегмент линий доступа расположенный между УД и оборудованием пользователя обозначен – ЛС. Предложены расчеты для следующих вариантов подключения пользователей:

- при создании СД не применяются УД и все пользователи подключены непосредственно к УК;
- одна часть пользователей, в пределах создаваемой СД, подключена к УК, а другая часть – к УД;
- пользователи, расположенные в пределах территории, обслуживаемой УК, подключены к УК, часть пользователей, расположенных в пределах ТУД, также подключены к УК, остальные пользователи ТУД включены в УД.

Для третьего варианта подключения пользователей к УК предложены формулы для расчета сегмента ЛД. Для модели прямоугольной структуры СД [3]:

$$\begin{aligned}
 I_{\text{пр}} = & \frac{\int_{-b_{\text{УК}}/2}^{b_{\text{УК}}/2} \eta a_{\text{УК}} y dy + \int_{-a_{\text{УК}}/2}^{a_{\text{УК}}/2} \eta b_{\text{УК}} x dx}{\eta a_{\text{УК}} b_{\text{УК}}} \frac{N_{\text{УК}}}{N_{\text{ТУК}} + N_{\text{УК}} + N_{\text{УД}}} + \frac{\int_{-b_{\text{УД}}/2}^{b_{\text{УД}}/2} \eta a_{\text{УД}} y dy + \int_{-a_{\text{УД}}/2}^{a_{\text{УД}}/2} \eta b_{\text{УД}} x dx}{\eta a_{\text{УД}} b_{\text{УД}}} \frac{N_{\text{УД}}}{N_{\text{ТУК}} + N_{\text{УК}} + N_{\text{УД}}} \\
 & + \frac{\left(\int_{-b_{\text{ТУК}}/2}^{b_{\text{ТУК}}/2} a_{\text{ТУК}} y \eta dy + \int_{-a_{\text{ТУК}}/2}^{a_{\text{ТУК}}/2} b_{\text{ТУК}} x \eta dx \right) - \left(\int_{-b_{\text{УК}}/2}^{b_{\text{УК}}/2} a_{\text{УК}} y \eta dy + \int_{-a_{\text{УК}}/2}^{a_{\text{УК}}/2} b_{\text{УК}} x \eta dx \right)}{\eta (ab - a_{\text{УК}} b_{\text{УК}})} \frac{N_{\text{ТУК}}}{N_{\text{ТУК}} + N_{\text{УК}} + N_{\text{УД}}} \\
 & + \frac{\left(\int_{-b_{\text{ТУК}}/2}^{b_{\text{ТУК}}/2} a_{\text{ТУК}} y \eta_{\text{УД}} dy + \int_{-a_{\text{ТУК}}/2}^{a_{\text{ТУК}}/2} b_{\text{ТУК}} x \eta_{\text{УД}} dx \right) - \left(\int_{-b_{\text{УК}}/2}^{b_{\text{УК}}/2} a_{\text{УК}} y \eta_{\text{УД}} dy + \int_{-a_{\text{УК}}/2}^{a_{\text{УК}}/2} b_{\text{УК}} x \eta_{\text{УД}} dx \right)}{\eta_{\text{УД}} (ab - a_{\text{УК}} b_{\text{УК}})} \frac{N_{\text{УД}}}{N_{\text{ТУК}} + N_{\text{УК}} + N_{\text{УД}}}
 \end{aligned} \tag{1}$$

где: $a_{\text{УК}}, b_{\text{УК}}$ – размеры сторон ТУК прямоугольной территории, где ЛД включены непосредственно в УК;

$a_{\text{УД}}, b_{\text{УД}}$ – размеры сторон ТУД;

η – поверхностная плотность распределения пользователей;

$\eta_{\text{УД}}$ – поверхностная плотность распределения УД;

$N_{\text{УД}}$ – количество ЛД, включенных в УД;

$N_{\text{УК}}$ – количество ЛД, включенных в УК;

$N_{\text{ТУК}}$ – количество ЛД, расположенных на территории, обслуживаемой УД, но включенных непосредственно в УК.

Для модели секторной структуры СД [4]:

$$\begin{aligned}
 I_{\text{сект}} = & \frac{\int_0^{R_{\text{ТУК}}} 2\pi\eta r^2 \alpha_{k_ТУК} dr}{\pi\eta R_{\text{УК}}^2} \frac{N_{\text{ТУК}}}{N_{\text{ТУК}} + N_{\text{УК}} + N_{\text{УД}}} + \frac{\int_{R_{\text{УК}}}^{R_{\text{ТУК}}} 2\pi\eta r^2 \alpha_{k_ТУК} dr}{\eta(\pi R_{\text{ТУК}}^2 - \pi R_{\text{УК}}^2)} \frac{N_{\text{УК}}}{N_{\text{ТУК}} + N_{\text{УК}} + N_{\text{УД}}} + \\
 & + \frac{\int_{R_{\text{УК}}}^{R_{\text{ТУК}}} 2\pi\eta_{\text{УД}} r^2 \alpha_{k_УД-УК} dr}{\eta_{\text{УД}}(\pi R_{\text{ТУК}}^2 - \pi R_{\text{УК}}^2)} \frac{N_{\text{УД}}}{N_{\text{ТУК}} + N_{\text{УК}} + N_{\text{УД}}} + \frac{\int_0^{R_{\text{ТУД}}} 2\pi\eta r^2 \alpha_{k_УД} dr}{\pi\eta R_{\text{УД}}^2} \frac{N_{\text{УД}}}{N_{\text{ТУК}} + N_{\text{УК}} + N_{\text{УД}}}
 \end{aligned} \tag{2}$$

где: η – поверхностная плотность распределения пользователей сети;

$\eta_{\text{УД}}$ – поверхностная плотность распределения УД;

$R_{\text{ТУК}}$ – радиус территории, обслуживаемой одним УК;

$R_{\text{УК}}$ – радиус территории, на которой ЛД включены непосредственно в УК;

$\alpha_{k_ТУК}$ – коэффициенты кривизны прокладки ЛД, включенных непосредственно в УК;

$\alpha_{k_УД-УК}$ – коэффициенты кривизны линий транспортного доступа между УД и УК;

$\alpha_{k_УД}$ – коэффициенты кривизны прокладки ЛД, включенных в УД;

$N_{\text{УД}}$ – количество ЛД, включенных в УД;

$N_{\text{УК}}$ – количество ЛД, включенных в УК;

$M_{\text{тук}}$ – количество ЛД, расположенных на территории, обслуживаемой УД, но включенных непосредственно в УК.

Операция интегрирования в (1) и (2) (вместо суммирования) позволяет учесть реальные пространственные характеристики трассирования в конкретных условиях прокладки линий.

Схема-алгоритм и программная реализация модели

На основе предложенной математической модели разработан алгоритм автоматизированного расчета характеристик СД, представленный на рисунке 3.

Главная проблема проектирования СД заключается в определении оптимального места расположения УК, мест расположения УД, в установлении границ территорий, обслуживаемых этими УД и оптимизации направлений прокладки ЛД. Проблема определения места расположения пользователей, т. е. вопрос, где и в каком количестве они появятся на территории СД, имеет существенно большее значение для выбора структуры СД, чем ответ на вопрос, в какой момент будут задействованы терминалы пользователей.

Информация о поверхностной плотности пользователей, их географическом размещении на рассматриваемой территории, интенсивности нагрузки загружается из внешнего файла. В зависимости от средней поверхностной плотности пользователей определяется место размещение УК. Территория, обслуживаемая СД, разбивается на сегменты, обслуживаемые УД, которые в свою очередь подключаются к УК. При моделировании используются арифметический и геометрический методы определения мест расположения УД, а так же итеративное определение мест расположения УД и границ ТУД. Применяя арифметический метод, расположение УД определяется путем оценки географического размещения пользователей и их количества. Этот метод выгодно применять, если между местами расположения пользователей и УД существуют однозначные линейные зависимости. При применении геометрического метода расположение УД определяется путем минимизации затрат на сегмент ЛД.

Для осуществления анализа влияния вариаций параметров сети на ее характеристики разработано программное обеспечение, которое реализует модель сети доступа. Программа позволяет загрузить карту распределения абонентов на местности из внешнего файла и изменять параметры сети для осуществления дальнейшего анализа. Программа представляет собой совокупность модулей и открыта для доработки и модернизации.

Заключение

Задачи проектирования и планирования сетей доступа вытекают из необходимости заблаговременного выбора технических средств, обеспечивающих удовлетворение потребностей пользователей в доступе к ИКУ и в обмене информацией. При внедрении научных методов в область проектирования в первую очередь возникает потребность в достоверных прогнозах. Правильный прогноз позволяет изначально выбрать соответствующую структуру сети, так как любые позднейшие изменения связаны с дополнительными расходами. Так как ошибка прогнозирования может привести к существенным незапланированным затратам, определение влияния таких ошибок на характеристики сети является важной задачей и заслуживает рассмотрения. В работе был предложен метод оценки чувствительности сетей доступа к вариации прогнозируемых входных параметров при использовании прямоугольной и радиальной моделей СД. В настоящее время, на базе предложенного подхода, ведется работа по имитационному и статистическому моделированию процесса создания СД и анализируются полученные результаты, которые планируется представить в следующей работе.



Рис. 3. Схема-алгоритм процедуры расчета характеристик СД

Литература

Гайворонская Г.С. Основные задачи модернизации сетей пользовательского доступа / Гайворонская Г.С., Котова А.И. // Зв'язок. – 2010. – №1 (89). –С. 18-24;

Гайворонская Г.С. Дослідження впливу помилок прогнозу вихідних даних на процес планування мереж доступу / Гайворонська Г.С., Сахарова С.В. // Збірник наукових праць ВІТІ НТУУ „КПІ”. – 2010. – № 2., с 23-29.

- Гайворонская Г.С. Метод определения местоположения узлов при использовании прямоугольной модели сети доступа / Г. С. Гайворонская, С.В. Сахарова // Холодильная техника и технология. – Одесса : ВЦ ОГАХ, 2011. (в печати)
- Гайворонская Г.С. Особенности определения местоположения узлов доступа при использовании радиальной модели обслуживаемой территории / Г. С. Гайворонская, С.В. Сахарова // Наукові праці ДонНТУ. – Донецьк, 2011 (в печати)
- Гайворонская Г.С. Концепция пользовательского доступа: Учебник для ВУЗов. – Одесса: ОГАХ, 2008. – 408 с.
- Соколов Н. А. Сети абонентского доступа. – Пермь: ИПК Звезда, 1999. – 154 с.
- Гайворонская Г.С. Оценка влияния некоторых факторов на процесс развития телекоммуникационных сетей // Холодильна техніка і технологія. – 2006. – №2 (100). – С. 95-100.
- Сахарова С.В. Задача выбора параметров сети доступа / С.В. Сахарова // Сборник тезисов Третьей международной конференции «Проблеми телекомунікацій», КПІ, Київ, 21-24 апреля 2009;
- Гайворонская Г.С. Классификация параметров сетей доступа / Г.С. Гайворонская, С.В. Сахарова // Сборник тезисов Пятой международной НТК «Современные информационно-коммуникационные технологии», Крым, Ялта, Ливадия, 05-09 октября 2009;
- Гайворонская Г.С. Анализ влияния вариации исходных параметров на результаты сетевого планирования // Тр. УНИИРТ. – 2006. – №3 (47). – С. 102-106;

Благодарности

Настоящая работа была выполнена при поддержке интернационального проекта ITHEA XXI Института информационных теорий и их приложений FOI ITHEA и Ассоциации ADUIS Украина (Ассоциация разработчиков и пользователей интеллектуальных систем)., в связи с чем автор и выражает свою искреннюю благодарность.

The paper is partially financed by the project ITHEA XXI of the Institute of Information Theories and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua). www.ithea.org , www.aduis.com.ua

Информация об авторах

Светлана Сахарова – *Институт информационных технологий Одесской государственной Академии холода, аспирант кафедры информационно-коммуникационных технологий; ул. Дворянская, 1/3, Одесса-26, 65026, Украина; тел. (048)-7209148; моб. (38067)-483-39-47; E – m: switchonline@rambler.ru*

Главные области научного исследования: Проблемы создания перспективных сетей доступа

ВОЗМОЖНЫЙ ПОДХОД К РАЗРАБОТКЕ МОДЕЛИ ТРАФИКА КОНВЕРГЕНТНОЙ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ

Максим Соломицкий

Аннотация: Предложено формальное описание конвергентной телекоммуникационной сети и процессов в ней. Выполнен анализ возможности использования математического аппарата теории телетрафика для создания модели трафика конвергентной телекоммуникационной сети. Определена необходимость разработки модели взаимодействия конвергентной телекоммуникационной сети с внешней средой с учетом самоподобного характера поступающего потока.

Ключевые слова: конвергентная телекоммуникационная сеть, поток вызовов, математическая модель.

Ключевые слова классификации ACM: G.3 PROBABILITY AND STATISTICS – Distribution functions, G. Mathematics of Computing – G.0 GENERAL, G.3 PROBABILITY AND STATISTICS – Stochastic processes.

*Для мага самая трезвая вещь, которую он может сделать,
– это перевести эти сущности на абстрактный уровень.
Чем меньше интерпретаций делают маги, тем лучше.*

К. Кастанеда

Введение

Современное общество производит и потребляет разные виды информации, для доставки которой используются различные информационные и телекоммуникационные сети. Такое разнообразие сетей является технически нецелесообразным и экономически невыгодным. В связи с этим особую актуальность приобретает проведение фундаментальных и прикладных научно-исследовательских работ в области создания современных телекоммуникационных сетей, обслуживающих разнородный трафик.

При решении задач проектирования и оптимизации телекоммуникационной сети (ТС) актуальной остается задача определения нагрузки поступающей на сеть. Это связано с тем, что величина нагрузки является одним из основополагающих факторов при расчете необходимого количества каналов связи, систем коммутации и другого сетевого оборудования. Существующие методы расчета сетевого оборудования основаны на моделях нагрузки, разработанных для таких известных классов сетей, как информационно-вычислительные [1,2], передачи данных [3,4], телефонные, которые призваны реализовать обслуживание пользователей только в определенной четко ограниченной области. Создаваемую в настоящее время конвергентную телекоммуникационную сеть (КТС) нельзя отнести ни к одному из этих классов. Поэтому существующие модели взаимодействия ТС с внешней средой (модели нагрузки) и методы расчета сетевого оборудования непригодны для решения задач анализа и проектирования КТС.

В связи с этим, на первый план выходят задачи развития и модернизации моделей и методов описания архитектуры ТС и процессов в этих сетях, которые можно рассматривать в качестве основы для создания математического аппарата разработки и усовершенствования моделей и методов синтеза и модернизации КТС. При этом представляет интерес исследование режимов работы ТС в процессе их

реконструкции и модернизации, когда существенно изменяется интенсивность внешнего воздействия на сетевые элементы. Структурная сложность КТС и ее отдельных частей, большое количество и разнообразие исходных данных, необходимость прогнозирования этих данных – все это обуславливает высокий уровень теоретической сложности решения задач построения и развития сети, в особенности, касающихся исследования ее динамики. Следовательно, нужны методы, позволяющие создать аппарат расчета сети в условиях ее динамического развития [5].

В статье под КТС понимается совокупность аппаратно-программных средств и архитектурно-технологических методов доставки информации территориально удаленным пользователям, позволяющая на единой цифровой основе обеспечить различные виды услуг по обработке и доставке разнородной информации, при обеспечении требований пользователей к своевременности и качеству доставки циркулирующей в сети интегральной информации. Объектом исследования данной работы является взаимодействие КТС с внешней средой. Согласно приведенному определению КТС накладывает ограничения на предмет исследования – потоки информации, формирующие нагрузку на сеть.

1. Конвергентная телекоммуникационная сеть как объект исследования и проектирования

Развитие концепции конвергентных телекоммуникационных сетей обусловлено, главным образом, необходимостью предоставления всего спектра современных информационно-коммуникационных услуг (ИКУ) в рамках единой универсальной ТС, обладающей технической возможностью организации новых служб для обеспечения рентабельности сети в будущем.

При планировании, построении, эксплуатации и развитии КТС возникает совокупность проблем, в каждой из которых, и всей их совокупности, можно выделить задачи, требующие взаимной увязки и совместимых оптимизационных решений. Согласно [6], первая из этих задач – проведение анализа потребностей пользователей в услугах связи, для удовлетворения которых создается сеть, – т.е. определение ожидаемой нагрузки на сеть. Такой анализ включает определение потенциальных объемов информационных потоков, распределение их в пространстве и времени. Здесь планирование заключается в уточнении видов и объемов традиционных и новых услуг связи, выборе правильных соотношений между объемами агрегируемых в потоке трафика различных ИКУ и определении характера распределения потоков, созданных этими услугами. Вторая задача планирования развития – оптимизация структуры КТС и распределения в ней информационных потоков. Третья задача, возникающая уже на фазе эксплуатации КТС, – это экспериментальное исследование такой системы.

Рассматривая КТС как большую систему, необходимо на этапах проектирования провести декомпозицию структуры и выполняемых ею функций на отдельные звенья и компоненты. При этом, исходя из целевого назначения КТС, выделим в качестве базового процесса, составляющего суть функционирования системы, процесс преобразования информации, т. е. информационную технологию. Такой подход позволит с единой системной позиции рассмотреть все составляющие технологического процесса преобразования информации в КТС на различных уровнях абстракции и дать общую характеристику процесса на уровне системы в целом [6].

Исходя из динамики процесса преобразования информации в КТС, выделим две его составляющие: вычислительный процесс, отражающий преобразование информации в основном во времени, и информационный процесс, связанный с преобразованием информации в основном в пространстве. Очевидно, что такое деление достаточно условно, как впрочем, и всякая классификация. Вычислительный и информационный процессы, как компоненты единого технологического процесса преобразования информации, находятся в тесной взаимосвязи, дополняя друг друга в зависимости от целей исследования и уровня рассмотрения. Вычислительный процесс состоит из подпроцессов таких как: ввод, вывод, хранение, обработка данных и т. п., а информационный процесс включает сбор, распределение, передачу, коммутацию информационных потоков и т. п. Остальные подпроцессы в сети учтем как

внешние воздействия. Проведение декомпозиции технологического процесса преобразования информации в КТС позволит оценить вероятностно-временные характеристики вычислительного и информационного процессов на уровне разработки структурных элементов системы, а также вероятностно-временные характеристики их составляющих. Это важно, поскольку объем необходимого оборудования зависит не только от интенсивности потока и вероятности потерь вызовов, но и от вероятностно-временной структуры потока вызовов. Таким образом, для правильного построения и расчета сетей необходимо иметь максимально точную информацию о поступающей нагрузке и о характере потока вызовов, создающего эту нагрузку [7]. Так как решение задачи построения и оптимизации КТС с системных позиций должно быть получено с учетом тесного взаимодействия сети с пользователями, необходимо ввести понятие входящего информационного потока, отражающего и позволяющего описать воздействие пользователей на КТС.

2. Формальное описание сетевого трафика и процессов, отражающих воздействие пользователей на КТС

В работе под сообщением пользователя (СП), т. е. информационным сообщением, поступающим из внешней информационной метаструктуры в транспортную систему КТС, будем понимать основную исследуемую единицу информации пользователя, т. е., по терминологии информационно-вычислительных сетей, «конечную последовательность данных, формируемую для передачи и имеющую законченное смысловое значение» [8]. В сети СП передаются в виде коммутируемых информационных единиц (КЕ), например сообщений, пакетов, дейтаграмм и т. д., длина которых может отличаться или не отличаться от длины СП. Трафик в сети может быть представлен в виде потоков с интенсивностью λ , причем для сетевых данных $\lambda = \lambda(\alpha)$, где α – в общем случае вектор, составляющие которого определяются алгоритмами управления обменом и коммутацией в КТС, а также структурой сети и ее узлов коммутации (УК). В КТС от пользователей поступают различные виды информации (оперативные и диалоговые данные, речь, видеопотоки, фоновая информация и т. д.), которые формируют агрегированный трафик КТС. Доля каждой из агрегируемых информационных категорий в общем объеме информации в КТС и в отдельных ее участках может меняться в течение времени в широких пределах.

В соответствии с режимами коммутации, приоритетами и видами информации различные виды поступающих от пользователей запросов на ресурсы КТС можно классифицировать в виде трехмерной матрицы, элементы a_{ij}^k которой являются условными (закодированными) обозначениями запросов, т. е.

a_{ij}^k – код запроса i -го вида информации j -го приоритета на k -й режим коммутации. Каждый запрос, кроме кода, имеет характеристики (параметры) p_0^D , T^D , D^D , которые могут быть фиксированы для каждого отдельного запроса, где p_0^D – максимальная допустимая вероятность искажения символа; T^D – максимально допустимая средняя задержка СП или КЕ; D^D – допустимая дисперсия времени задержки.

Для полноты морфологического описания КТС необходимо описание модели сетевого трафика. Модель трафика определяется законом распределения интервалов времени между моментами поступлений СП; законом распределения числа КЕ в СП; классом приоритетности, определяемым в общем случае срочностью доставки; ценностью информации; адресом источника и адресом получателя информации и т. д. В связи с тем, что основным назначением КТС является обеспечение требуемого качества обслуживания пользователей (реализация пользовательских требований на передачу информации), поток поступивших в сеть СП определяет нагрузку на КТС.

3. Формальное описание архитектуры КТС

Рассматривая взаимодействие внешней среды и элементов КТС и обработку цифровой информации каждым из элементов этой сети в соответствии с принципами модели взаимодействия открытых систем (ВОС), проведем формальное описание архитектуры КТС в виде внешней среды и сетевой среды. Каждая из этих сред представляется в свою очередь в виде двух сред: внешняя среда – средами генерации и распространения, сетевая среда – средами взаимодействия и обработки. Предложенная иерархия сред отображена на рис.1.

Выделение среды генерации позволит описать воздействие пользователей на сеть, т. е, потоков их запросов на ресурсы сети; среда распространения позволяет отразить воздействие внешних мешающих факторов (помех, сбоев и т. п.) на распространение физических сигналов (электрических, оптических и т.д.); среда взаимодействия дает возможность описать сеть в целом, как единую систему, взаимодействующую с внешней средой на системном уровне (уровне сред генерации и взаимодействия); среда обработки – возможность описать основные структурные элементы сети, обрабатывающие цифровую формацию: УК, концентраторы, абонентские пункты, тракты передачи и т. п. с учетом взаимодействия с внешней средой на реализационном (аппаратно-программном) уровне (уровне сред распространения и обработки).

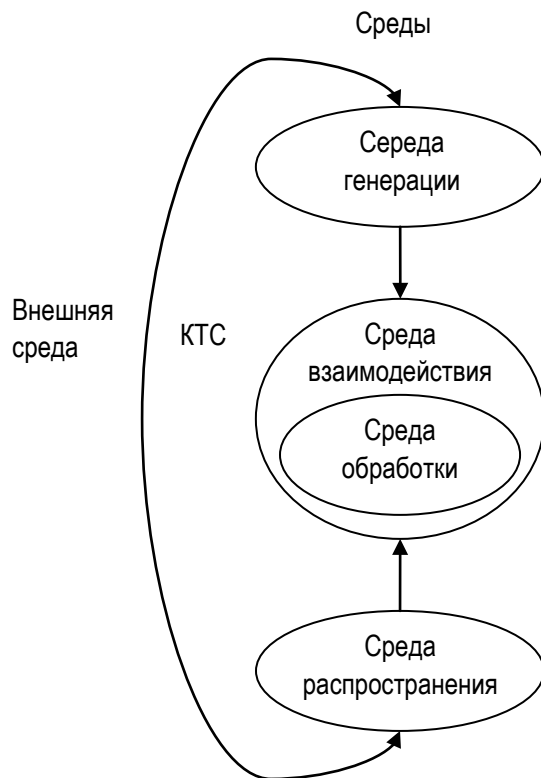


Рисунок 1 – Иерархия сред согласно формальному описанию КТС

Среду взаимодействия представим как некоторую однородную или регулярную информационно-вычислительную среду [9], состоящую из УК. Рассматриваемые узлы представляется целесообразным формализовать в виде совокупности открытых систем. Согласно модели ВОС все соединения между открытыми системами являются дуплексными, а рассматриваемые узлы среды взаимодействия подчиняются следующим правилам:

- связь между объектами одного и того же уровня двух или более открытых систем в рамках узла среды взаимодействия может осуществляться через объекты смежных уровней, исключая физический;
- объекты открытых систем, находящиеся в разных узлах среды взаимодействия, соединяются между собой обязательно через физические уровни этих систем.

Для описания и анализа особенностей среды взаимодействия можно воспользоваться предложенным в [9] аппаратом иерархических матриц связности. Структура любой сети может быть задана с помощью простой квадратной матрицы связности (МС), размерность которой определяется числом узлов в сети N_u . Однако такая МС будет иметь большую размерность при применении в КТС, к тому же в ней в явном виде нельзя отразить факт наличия и особенности некоторых вторичных сетей. Аппарат же иерархических матриц, представляющий собой совокупность квадратных и прямоугольных МС, позволяет устранить эти недостатки.

Среду обработки (информационно-вычислительную среду) представим в виде децентрализованных структур однотипных элементов – абстрактных узлов и коммутационных систем, показанных на рис. 2 [9,10]. Понятие «абстрактного узла» взято из концепции «теории разговоров», в рамках которой это понятие использовалось для формального описания протоколов в сетях передачи данных и вычислительных сетях. В силу того, что «теория разговоров» характеризуется, прежде всего, тем, что она строится на формулировке связей и правил обмена между соседними узлами сети, при этом более сложные структуры создаются путем последовательного наращивания, т. е. на одном уровне, понятие «абстрактного узла» в исходной форме представляется неприемлемым, т. к. подобные структуры трудно реализуются при наличии нескольких одновременных «разговоров» из-за ограничений по производительности, и, следовательно, для реализации УК с необходимой для КТС высокой производительностью понятие «абстрактного узла» необходимо переработать. При этом предположим, что полученное формальное представление процессов в сетевых узлах на базе среды обработки, описываемой в виде абстрактных узлов с приемлемой для КТС структурой, позволит решать задачи идентификации процессов коммутации и обмена информацией на различных уровнях иерархии КТС.

Выделение среды распространения отражает наличие стохастических мешающих воздействий внешней среды на распространение информации в КТС, что соответствует 1 и 2 уровням модели ВОС. Наиболее простой из математических моделей для описания среды распространения является модель независимых ошибок, которая исходит из предположения, что ошибки в дискретном канале связи возникают независимо друг от друга, т. е. между ними отсутствует корреляция.

Экспериментальные работы в области исследования каналов связи не подтвердили гипотезы о независимости ошибок, а выявили тенденцию к их группированию (пакетированию) [11].

Известно много математических моделей дискретных каналов связи, в которых различным образом отражается экспериментально установленный факт группирования ошибок. Существующие модели по способу описания параметров потоков ошибок можно условно разделить на две большие группы. Модели первой группы отражают чисто математический подход к вопросу описания потока ошибок, но при этом, как правило, игнорируется физическая сторона процессов, происходящих в каналах связи, и механизм группирования ошибок не задается в явном виде. В моделях второй группы делаются попытки в какой-то степени учесть физические явления, которые приводят к появлению ошибок. В моделях этой группы используется понятие пакета ошибок, механизм образования которых задается в явном виде.

Наиболее целесообразным представляется строить математическую модель среды распространения, оперируя понятиями, близкими к физическим явлениям, происходящим в каналах связи, т. е. на 1 уровне модели ВОС.

Разработка адекватной модели для описания ошибок в каналах связи различных типов, используемых в КТС, является целью отдельного глубокого исследования и в рамках данной работы не рассматривается.

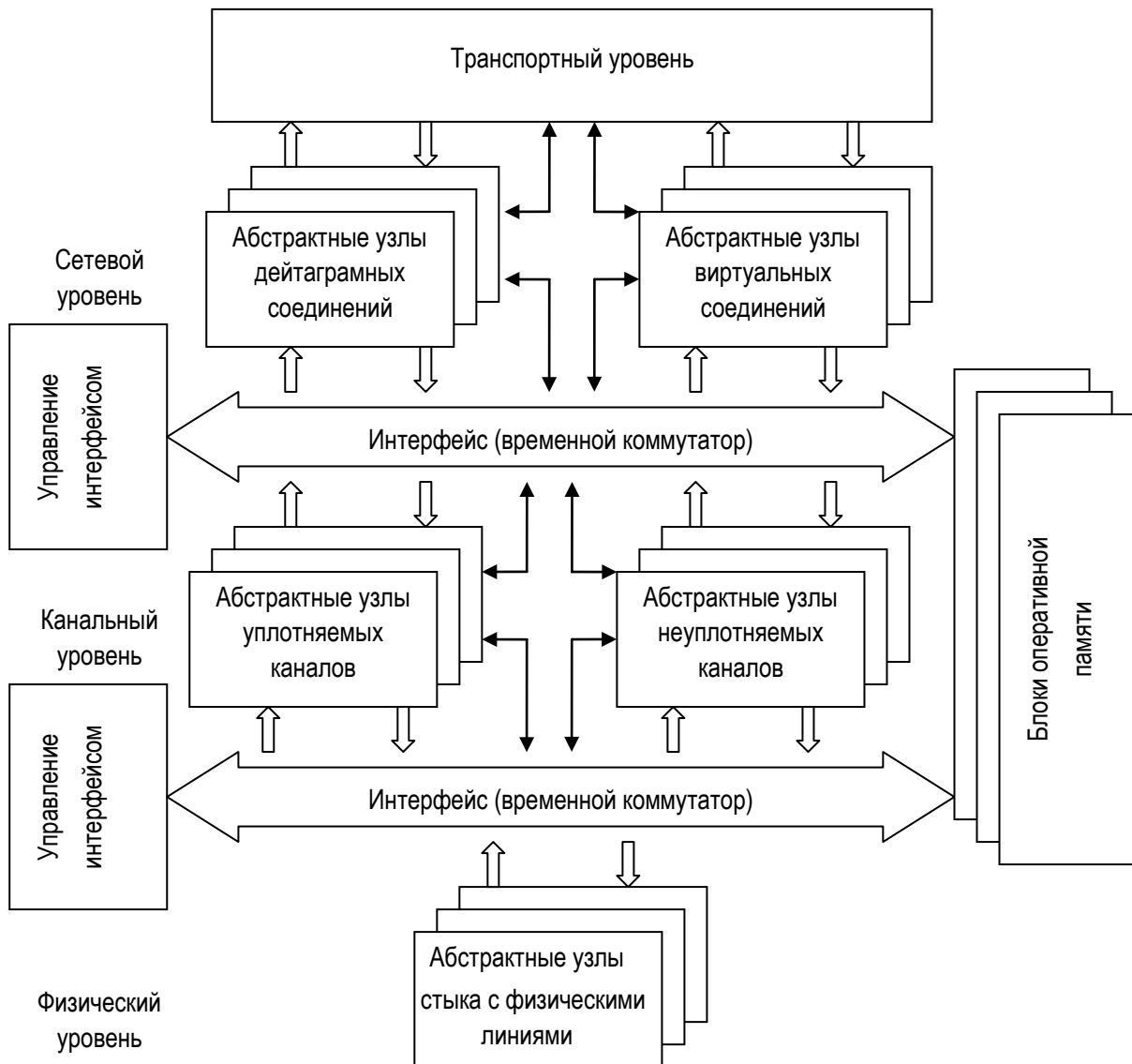


Рисунок 2 – Формальное представление среды обработки

4. Модель взаимодействия КТС с внешней средой как математическая основа модели нагрузки КТС

Согласно формальному описанию, введенному в предыдущем разделе, внешние воздействия на КТС со стороны пользователей представляются в виде среды генерации, отражающей динамику пользовательских запросов на ресурсы КТС и характер этих запросов. С точки зрения специалистов, занимающихся исследованием и проектированием системы управления КТС, любая модель взаимодействия с внешней средой должна рассматриваться как математическая основа, позволяющая создать приемлемые для практики методы идентификации и анализа вероятностно-временных характеристик процесса функционирования сети. Рассмотрение передачи различных видов информационных сообщений пользователей (СП) в виде потоков коммутируемых информационных

единиц (KE) дает возможность перейти к формальному описанию среды генерации, сводимому к описанию воздействия входящих потоков (вызовов) на сеть.

Модель потока вызовов изучается достаточно давно. В [12] потоком вызовов (событий) называется последовательность вызовов, поступающих через какие-либо интервалы или в какие-либо моменты времени. При этом в опубликованных работах поток вызовов определяется тремя эквивалентными способами: последовательностью вызывающих моментов, последовательностью промежутков времени между вызывающими моментами и последовательностью чисел, определяющих количество вызовов, поступающих в течение заданных отрезков времени. Такие способы задания потоков являются исчерпывающим при рассмотрении передачи речи по телефонной сети с позиционным мультиплексированием. Однако КТС обслуживает трафик с использованием как позиционного, так и меточного мультиплексирования. Поэтому очевидно, что поток КТС помимо распределения во времени должен определяться и количеством передаваемой информации.

Выполним анализ основных свойств потоков, поступающих на ТС, на основании базисных работ [12–14] и определим возможность применения рассмотренных потоков для КТС.

Вероятность поступления некоторого количества требований в промежутке определенной длины для стационарного ординарного потока без последействия, называемого простейшим потоком, рассчитывают по формуле Пуассона. При этом распределение вероятностей промежутка между поступлением соседних требований подчиняется экспоненциальному закону, а математическое ожидание этой, экспоненциально распределенной, случайной величины равно ее дисперсии. Для применения этой математической модели на КТС необходимо наложить большое количество ограничений и проводить ее исследования с детерминированными режимами использования сетевых ресурсов, которые бы позволили удовлетворить предположениям о стационарности и ординарности поступающего в сеть потока. Однако даже в таком случае модель не может быть пригодной, т.к. предположение об отсутствии последействия потока не соответствует действительности, поскольку, как это будет показано ниже, поток КТС обладает долгосрочной зависимостью поступления информации. В виду этого представляется целесообразным использование свертки вероятностей поступления требований по формуле Пуассона и аппарата для определения корреляции вероятностного процесса поступления требований. Это же касается и одного из важных свойств пуассоновского потока – аддитивности.

Для применения математических моделей нестационарного (ординарного потока без последействия) и неординарного (группового стационарного потока без последействия) пуассоновских потоков к КТС необходимо, как и в случае простейшего потока, учесть и определенным образом математически описать корреляции между поступлениями требований. Представляется целесообразным переработать и модернизировать параметры указанных нестационарного и неординарного потоков, что позволит определить наличие и задать характер зависимости между моментами поступления требований в потоках. Это же касается и группы нестационарных потоков (симметричного, примитивного и потока с повторными вызовами) с простым последействием.

Для применения моделей потоков (Пальма, Эрланга) с ограниченным последействием в качестве математической модели потока КТС необходимо пересмотреть характер, рассматриваемого в рамках этих моделей, последействия. Для этого необходимо задать математическое описание корреляции промежутков между требованиями. В частности, в случае потока Эрланга [15] характер изменений тождествен вышесказанному в рамках анализа пуассоновских потоков.

Подытожив можно сказать, что математический аппарат теории телетрафика [12–14] дает хорошие результаты при его применении для телефонных сетей с коммутацией каналов и иногда удовлетворительные результаты для сетей передачи данных, вычислительных систем с коммутацией пакетов и сообщений. Однако ни один из рассмотренных потоков не пригоден для адекватного описания потока интегрального трафика циркулирующего внутри КТС, полученного в результате развития, взаимопроникновения и слияния информационных и коммуникационных технологий на протяжении нескольких десятков лет [16].

Хорошо проработанные вопросы анализа математических моделей систем массового обслуживания (СМО) с ожиданием [17] имеют важное значение для проектировщиков сетей и системных аналитиков и очень полезны при проектировании, планировании ресурсов и предсказании производительности сети. Однако во многих реальных случаях результаты, полученные на основе анализа очередей в СМО, существенно отличаются от фактически наблюдаемой производительности, т.к. анализ очередей применим лишь в случае, когда трафик данных подчиняется распределению Пуассона [18].

Многочисленные исследования [20–27], проведенные за последнее десятилетие прошлого века и в начале нынешнего учеными разных стран, позволяют утверждать, что во многих ситуациях трафик современных сетей обладает особой структурой, не позволяющей использовать при проектировании привычные методы, основанные на Марковских моделях и формулах Эрланга, которые хорошо себя зарекомендовали как аппарат проектирования телефонных сетей с коммутацией каналов. Игнорирование этих особенностей трафика приводит к недооценке или переоценке нагрузки и к неоправданно оптимистическим решениям [19]. Особенности, о которых идет речь, вызваны проявлением эффекта самоподобия трафика.

В статье [20] показано, что анализ очередей с использованием предположения о пуассоновском потоке не представляет адекватную модель современного сетевого трафика. Опираясь на большое количество исходных данных и статистический анализ, авторы утверждают, что для трафика в сетях с применением технологии *Ethernet* требуется новый метод моделирования и анализа сетевого трафика. Эта статья вызвала поток исследовательских работ по данному вопросу. Анализ множества статистических тестов [20] приводит к выводу, что трафик сети, работающей по технологии *Ethernet*, является самоподобным с параметром Херста $H=0,9$. Характер трафика, полученного при помощи самоподобной модели с $H=0,9$, оказался довольно близким к данным, полученным в реальных условиях. Авторами показано, что метод, заключающийся в моделировании трафика сети с использованием технологии *Ethernet* путем суперпозиции нескольких Парето-подобных источников *ON/OFF*, каждый из которых находится в одном из двух состояний: *ON*, в котором он активно передает пакеты, и *OFF*, в течении которого он бездействует, во-первых, позволяет получить результаты, сходные с данными реального трафика сети, работающей по технологии *Ethernet*, во-вторых, для него требуется определить совсем немного параметров. Исследователями было обнаружено, что изучаемый трафик сети с применением технологии *Ethernet* соответствует самоподобному трафику с $H=0,9$ [21].

В [22] сообщается об исследовании трафика Интернет, включающем более полумиллиона запросов к *web*-документам. Данные, полученные по методологии сходной с той, что применялась в исследовании сети, работающей по технологии *Ethernet*, показали, что трафик, формируемый *web*-браузерами, является самоподобным. Сходные результаты опубликованы в [23,24]. Исследуемые данные хорошо соответствуют распределению Парето, в процессе исследований обнаружено, что файлы, доступные в Интернет, обладают медленно затухающими распределениями интенсивности объемов передаваемых данных при загрузке файлов пользователями.

Исследования [25] показали, что модели, основанные на пуассоновском распределении, не соответствуют поведению сетей с системой сигнализации № 7, при этом обнаружилось, что модели самоподобного трафика обеспечивают лучшее соответствие для потоков управляющих сигналов, формируемых в цифровых телекоммуникационных сетях (цифровой сети интегрального обслуживания и др.).

В [26] сообщается об исследовании широкого спектра потоков, циркулирующих в сетях, работающих по протоколу *Transmission Control Protocol (TCP)*, а также об анализе трафика в сетях, работающих по протоколам *File Transfer Protocol (FTP)* и *Terminal Network (TELNET)*. При этом сделаны следующие общие выводы.

В используемых обычно пуассоновских моделях существенно недооценивается неравномерность трафика сетей, работающих по протоколу *TCP*, в широком диапазоне временных шкал.

Для интерактивного трафика сетей, работающих по протоколу *TELNET*, процесс поступления вызовов с последующим установлением соединений хорошо моделируются пуассоновским распределением, однако предположение о пуассоновском распределении поступления пакетов существенно недооценивает неравномерность трафика.

Для групповой пересылки данных, осуществляемой по протоколу *FTP*, структура трафика заметно отличается от пуассоновского; процесс поступления заявок, приводящих к последующему установлению сеансов связи по протоколу *FTP*, хорошо соответствует пуассоновской модели, но скорость поступления данных по соединениям *FTP* оказывается существенно более неравномерной, кроме того, распределение количества байтов в каждом сеансе является медленно затухающим.

Так же долгосрочно зависимыми (обладающими медленно затухающими распределениями) являются задержки доставки пакетов по протоколу *User Datagram Protocol* в Интернет [27].

Модель взаимодействия КТС с внешней средой, учитывающая фактор самоподобия и отражающая воздействие на сеть пользователей, формирующих поступающую нагрузку, на данный момент в опубликованных источниках не представлена.

Заключение

Сформулированные в работе формальные представления и подход позволят специалистам в области исследования и проектирования сетей, используя в качестве конструктивного принципа метод восхождения от, предложенного в работе, абстрактного к конкретному, воспроизвести КТС как в целом, так и отдельные ее части, и создать приемлемые для практики методы идентификации и анализа функционирования КТС. Предложенное в работе формальное описание сред КТС и входных потоков позволит описать взаимодействие КТС с внешней средой на системном уровне, что особенно важно как для последующей идентификации сети, как объекта управления, так и для построения самообучающихся моделей при адаптивном управлении сетью. Взаимодействие КТС с внешней средой позволит определить модель трафика КТС, необходимую для повышения качества обслуживания пользователей, математическую основу метода расчета нагрузки КТС.

Проведенное исследование моделей взаимодействия КТС с внешней средой указывает на невозможность использования существующего математического аппарата для определения вероятностно-временных характеристик КТС, сетьюобразующим процессом которой является конвергенция и, следовательно, внутри которой обеспечивается доставка сложной по своей природе и структуре интегральной информации, формируемой объединенным трафиком меточного и позиционного мультиплексирования). Очевидно, что потоки КТС помимо распределения во времени должны характеризоваться длиной отдельных сообщений, и для их описания представляется целесообразным использовать суперпозицию наиболее адекватных потоков, описываемых рассмотренными математическими моделями, с учетом того, что формируемый этими потоками трафик является самоподобным, как это показано в [28]. В работе показана необходимость разработки модели нагрузки КТС как математической основы метода расчета нагрузки, учитывающей самоподобный характер трафика в этой сети.

Благодарности

Настоящая работа была выполнена при поддержке интернационального проекта *ITHEA XXI* Института информационных теорий и их приложений *FOI ITHEA* и Ассоциации *ADUIS* Украина (Ассоциация разработчиков и пользователей интеллектуальных систем).

Литература

1. Бутрименко А. В. Разработка и эксплуатация сетей ЭВМ / Бутрименко А. В. – М.: Финансы и статистика, 1981. – 256 с.
2. Вычислительные сети (адаптивность, помехоустойчивость, надежность) / С. И. Самойленко, А. А. Давыдов, В. В. Золотарев и др. – М.: Наука, 1981. – 278 с.

3. Захаров Г. П. Методы исследования сетей передачи данных / Захаров Г. П. – М.: Радио и связь, 1982. – 208 с.
4. Клейнрок Л. Коммуникационные сети (стохастические потоки и задержки сообщений) / Клейнрок Л. – М.: Наука, 1970. – 256 с.
5. Гайворонская Г. С. Экономические аспекты модернизации телекоммуникационной сети / Гайворонская Г. С. // Зв'язок. – 2007. – №2 (70). – С.15-21.
6. Советов Б. Я. Построение сетей интегрального обслуживания / Советов Б. Я., Яковлев С. А.– Л.: Машиностроение, 1990. – 332 с.
7. Унификация метода расчёта оборудования телекоммуникационных сетей / Гайворонская Г. С., Сомсиков Д. А. и др. // Холодильна техніка і технологія. – 2008. – №5 (115). – С.65-68.
8. Вычислительные сети и сетевые протоколы / Д. Дэвис, Д. Барбер, У. Прайс и др. – М.: Мир, 1982. – 562 с.
9. Советов Б. Я. Разработка протоколов обмена интегральной информацией в информационно-вычислительных сетях с использованием рекомендации Х.25 // Рекомендация МККТТ Х.25 и ее применение в информационно-вычислительных сетях / Советов Б. Я., Цанов В. А., Яковлев С. А. – М.: изд. Междун. Центра науч.-техн. инф., 1983. – С. 83-103.
10. Разработка моделей для оценки эффективности методов управления в информационных сетях с учетом динамики их развития / В. П. Ильин, С. С. Йонев, С. А. Яковлев и др. // Системные принципы организации моделей развития. – М.: изд. НС «Кибернетика», 1981. – С. 41-52. – (Вопр. кибернетики; Вып. 72).
11. Gilbert E. Capacity of burst-noise channel // BSTJ / Gilbert E., – V. 39. – N. 3, – P. 941-957.
12. Лившиц Б.С. Теория телетрафика / Лившиц Б.С., Пшеничников А.П., Харкевич А.Д. – М.: Связь, 1979. – 224 с.
13. Элдин А. Основы теории телетрафика / Элдин А., Линд Г. – М.: Связь, 1972. – 200 с.
14. Крылов В. В. Теория телетрафика и ее приложения / Крылов В. В., Самохвалова С. С. – СПб.: БХВ-Петербург, 2005. – 288 с.
15. Гайворонская Г.С. О возможности применения метода этапов для имитационного моделирования потоков вызовов в сетях связи / Г.С. Гайворонская // Зв'язок. – 1998. – № 2. – С.35-36
16. Гайворонская Г. С. Инфокоммуникации: Учеб. пособие / Гайворонская Г.С. – Одесса: ОДАХ, 2006. – 90 с.
17. Клейнрок Л. Теория массового обслуживания / Клейнрок Л. – М.: Машиностроение, 1979. – 432 с.
18. В. Столлингс. Современные компьютерные сети/В. Столлингс.– СПб.: Питер, 2003.– 783с.
19. Телекоммуникационные системы и сети: Учебное пособие. В 3 томах. Том 3. – Мультисервисные сети /под ред. профессора В. П. Шувалова. – М.: Горячая линия-Телеком, 2005. – 592 с.
20. On the Self-Similar Nature of Ethernet Traffic (Extended version)/Will E. Leland, Murad S. Taqqu, Walter Willinger, Daniel V. Wilson// IEEE/ACM Transactions on Networking – 1994, February.
21. Self-Similarity Through High Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level / Willinger W., Taqqu M., Sherman R., and Wilson D. // IEEE/ACM Transactions on Networking – 1997, February.
22. Crovella M. Self-Similarity in World-Wide Web Traffic: Evidence and Possible Causes / Crovella M., Bestavros A. // Proceedings, ACM Sigmetrics Conference on Measurements and Modeling of Computer Systems – 1996, May.
23. Arlitt M. Workload Characterization of a Web Proxy in a Cable Modem Environment / Arlitt M., Friedrich R., Jin T. // ACM Performance Evaluation Review – 1999, September.
24. Arlitt M. A Workload Characterization Study of the 1998 World Cup Web Site / Arlitt M., Jin T. // IEEE Network – 2000, May/June.
25. Statistical Analysis of CCSN/SS7 Traffic Data from Working CCS Subnetworks / Diane E. Duffy, Allen A. McIntosh, Mark Rosenstein, Walter Willinger // IEEE Journal on Selected Areas in Communications – 1994, April.
26. Giroux N. Quality of Service in ATM Networks / Giroux N., Ganti S. // Upper Saddle River, NJ: Prentice Hall – 1999.
27. Borella M. Measurement and Analyses of Long-Range Packet Dependent Behavior of Internet Packet Delay / Borella M., Brewster G. // IEEE INFOCOM '98 – 1998, April.
28. Гайворонская Г.С. Исследование параметров объединенного потока вызовов / Г.С. Гайворонская // Труды УГАС "Информатика и связь". – 1997. – С. 222 – 226.

Информация об авторах

Максим Соломицкий – Одесская государственная Академия холода; магистрант кафедры информационно-коммуникационных технологий; ул. Дворянская, 1/3, Одесса – 82, 65082, Украина; e-mail: sage89@mail.ru

Главные области научных исследований: потоки вызовов, нагрузка в современных телекоммуникационных сетях. Проблемы создания конвергентных телекоммуникационных сетей

АНАЛИЗ ИНТЕНСИВНОСТИ НАГРУЗКИ НА РЕАЛЬНЫХ ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЯХ ЮГА УКРАИНЫ

Илья Ганницкий

Аннотация: Рассмотрено изменение нагрузки на современной телекоммуникационной сети. Приведены обобщенные результаты, а также определены общие тенденции развития данной сети.

Ключевые слова: телекоммуникационная сеть, параметры потоков вызовов, интенсивность нагрузки.

Ключевые слова классификации ACM: G.3 PROBABILITY AND STATISTICS – Distribution functions, G. Mathematics of Computing – G.0 GENERAL.

Введение

В последние годы на телекоммуникационных сетях (ТС) значительно увеличилась доля неречевой нагрузки. Это вызвано резким увеличением количества пользователей Интернет, электронного документооборота, увеличением электронных банковских операций, сверхвысокими темпами роста пользователей мобильной связи и т.п. Это и ряд других причин привели к существенному изменению параметров потоков вызовов на телекоммуникационных сетях Украины.

Однако после приобретения независимости в Украине практически (за небольшим исключением) не проводятся исследования потоков вызовов, циркулирующих в ТС.

В связи с этим исследования параметров потоков вызовов приобретают все более важное значение. Некоторым результатам исследования параметров потоков вызовов на ТС посвящена данная работа.

1. Анализ нагрузки на современной ТС

Исследования выполнялись на ТС крупного областного центра, являющегося одновременно важным транзитным узлом как крупного региона Украины, так и международного сообщения. Выбор данного областного центра обусловлен тем, что на протяжении многих лет этот город выявлял тенденции изменения потоков вызовов, которые со временем находили свое отражение и в других областных центрах и крупных городах Украины [1-4]. При этом выполнен анализ статистических данных параметров нагрузки за период 3,5 года (2003, 2004, 2005 и первое полугодие 2006 гг.), что составило 711608646 записей в базе данных, емкостью более 100 Гб.

Выбранные записи структурированы по двум параметрам: количество вызовов и длительность соединения.

В качестве примера выполненного анализа, рассмотрено изменение месячной нагрузки в марте 2003, 2004, 2005 и 2006 годов. Среднее и общее значения количества вызовов за данный месяц для различных типов нагрузки приведены в таблицах 1-3.

Рассмотрим соотношение нагрузки в пределах года. В 2003 году среднее значение интенсивности нагрузки для местных вызовов составило 127,96 Эрл; для междугородной нагрузки данное значение составило 122,74 Эрл; для операторов мобильной связи - 6,11 Эрл. В процентном соотношении данное значение составляет 49,8%, 47,8% и 2,4% соответственно. Максимальное значение интенсивности для местной нагрузки составляет – 250,7 Эрл, междугородной – 264,7 Эрл, операторов мобильной связи – 7,57 Эрл. Минимальное значение составило – 95 Эрл, 82,5 Эрл и 3,35 Эрл соответственно.

Таблица 1 – Местные вызовы

	2003 год	2004 год	2005 год	2006 год
Среднее количество вызовов за сутки	72000	179000 (увеличение на 249% по сравнению с 2003 г.)	247000 (увеличение на 138% по сравнению с 2004 г.)	213000 (87% от числа вызовов за 2005 г.)
Общее количество вызовов за март	2223011	5546105 (увеличение на 249% по сравнению с 2003 г.)	7648824 (увеличение на 138% по сравнению с 2004 г.)	6587008 (86% от количества вызовов за 2005 г.)

Таблица 2 – Междугородные вызовы

	2003 год	2004 год	2005 год	2006 год
Среднее число вызовов за сутки	41600	63200 (увеличение на 152% по сравнению с 2003 г.)	81900 (увеличение на 130% по сравнению с 2004 г.)	61400 (75% от количества вызовов за 2005 г.)
Общее количество вызовов за март	1289599	1958556 (увеличение на 152% по сравнению с 2003 г.)	2539441 (увеличение на 130% по сравнению с 2004 г.)	1904445 (75% от количества вызовов за 2005 г.)

Таблица 3 – Вызовы операторов мобильной связи, проходящие через телефонную сеть общего пользования

	2003 год	2004 год	2005 год	2006 год
Среднее количество вызовов за сутки	17800	91000 (увеличение на 511% по сравнению с 2003 г.)	167600 (увеличение на 184% по сравнению с 2004 г.)	158000 (94% от количества вызовов за 2005 г.)
Общее количество вызовов за март	553222	2822939 (увеличение на 510% по сравнению с 2003 г.)	5194522 (увеличение на 184% по сравнению с 2004 г.)	4902248 (94% от количества вызовов за 2005 г.)

В 2004 году наблюдается общее увеличение среднего значения интенсивности нагрузки как для местных вызовов – 283,06 Эрл (54,76 % от общего значения) и междугородних вызовов – 166,7 Эрл (32,24%), так и для вызовов от операторов мобильной связи – 67,15 Эрл (13%). По сравнению с 2003 годом интенсивность нагрузки возросла на 221% – для местных вызовов, на 136% для междугородних и практически на 1100% – для операторов мобильной связи. Максимальное значение интенсивности нагрузки для местных вызовов составило – 527,16 Эрл, для междугородней нагрузки – 410 Эрл и для операторов мобильной связи – 77,23 Эрл. Минимальное значение составило соответственно – 242 Эрл, 118,56 Эрл и 49,91 Эрл.

В 2005 году наблюдается дальнейшее увеличение интенсивности нагрузки. Для местных вызовов среднее значение составило 409 Эрл (53,15% от общей нагрузки); для междугородней нагрузки - 228,28 Эрл (29,67%); для операторов мобильной связи – 132,36 Эрл (17,18%). Темпы роста нагрузки сократились по сравнению с 2004 годом, и составили 144,5% для местных вызовов; 136,9% - для междугородних; 197,1% - для операторов мобильной связи. Максимальные значения интенсивности нагрузки – 901,01 Эрл, 718,33 Эрл и 156,88 Эрл для местной, междугородней и операторов мобильной связи соответственно. Минимальные значения – 304,33 Эрл, 181,37 Эрл и 92,97 Эрл.

В 2006 году среднее значение нагрузки для местных вызовов составило 377,47 Эрл; для междугородней нагрузки данное значение составило 177,38 Эрл; для операторов мобильной связи – 136,58 Эрл. В процентном соотношении данные значения составляют 54,6%, 25,6% и 19,8% соответственно. Интенсивность местной нагрузки в этом году составила 92,3% от интенсивности нагрузки за 2005 год, междугородняя – 77,7%, а нагрузка операторов мобильной связи составила 103,2%. Максимальное значение интенсивности для местной нагрузки составило – 924,70 Эрл, междугородней – 608,56 Эрл,

операторов мобильной связи – 155,59 Эрл. Минимальное значение составило – 315,66 Эрл, 139,25 Эрл и 89,94 Эрл соответственно. На рисунках 1-4 приведены графики изменения интенсивности нагрузки за март 2003-2006 гг.

В 2003 году наблюдается небольшое уменьшение количества вызовов в выходные дни для местных вызовов, однако пик интенсивности нагрузки приходится на воскресенье, а также наблюдается незначительное плавное повышение нагрузки к концу месяца. Для междугородней нагрузки характерно резкое увеличение интенсивности в выходные дни, что связано с льготным тарифом на этот период. Также наблюдается небольшое снижение нагрузки в выходные дни для операторов мобильной связи.

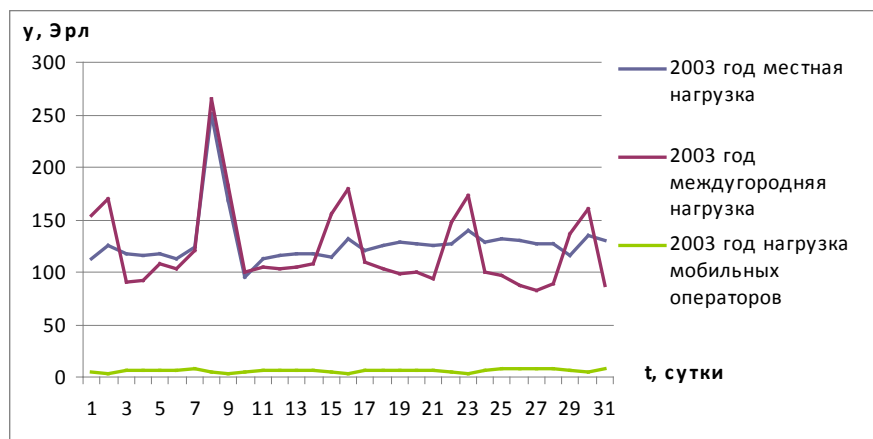


Рис. 1 График месячной нагрузки с суточным усреднением (март 2003 года).

В 2004 году наблюдается снижение доли междугородней нагрузки. При небольшом количестве вызовов в субботу и максимальном количестве вызовов в воскресенье, по сравнению с рабочими днями недели, наблюдается существенное увеличение интенсивности нагрузки в выходные дни. Для местных вызовов тенденции изменения нагрузки аналогичны 2003 году, т.е. наблюдается небольшое снижение нагрузки в субботу и пик нагрузки приходится на воскресенье. Для операторов мобильной связи кривая изменения нагрузки является слабо выраженной, однако наблюдается общее увеличение доли нагрузки, по сравнению с предыдущим годом, а также небольшой спад нагрузки в выходные дни.

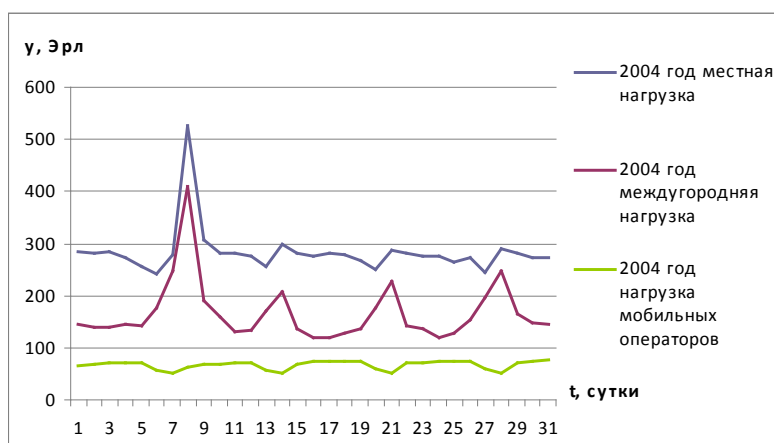


Рис. 2 График месячной нагрузки с суточным усреднением (март 2004 года).

В 2005 году при общем увеличении интенсивности нагрузки тенденции ее изменения аналогичны тенденциям 2004 года.

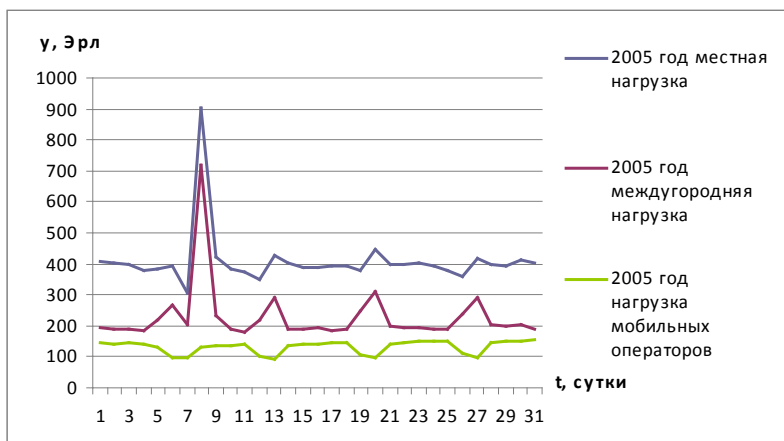


Рис. 3 График месячной нагрузки с суточным усреднением (март 2005 года).

В 2006 году наблюдается уравнивание доли междугородней нагрузки и нагрузки создаваемой операторами мобильной связи, что обуславливает необходимость более детального изучения вероятностно-временных характеристик создаваемых ими потоков вызовов. Общие тенденции изменения нагрузки сохраняются, т.е. для междугородней нагрузки характерно ее увеличение в выходные дни, а для нагрузки операторов мобильной связи – спад. Кривая, характеризующая местную нагрузку, становится все менее ярко выраженной и наблюдается незначительное повышение интенсивности нагрузки в воскресенье.

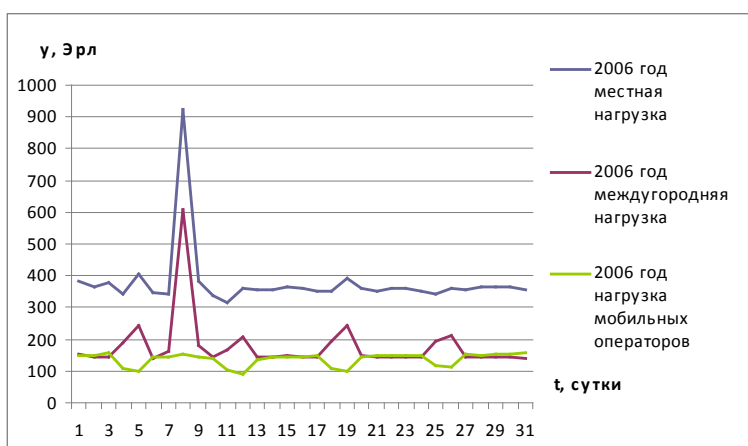


Рис. 4 График месячной нагрузки с суточным усреднением (март 2006 года).

Результаты измерений позволяют проанализировать изменение числа вызовов по дням месяца. При этом хорошо прослеживаются недельные колебания нагрузки и ярко выраженные пики, приходящиеся на праздник 8 Марта. Нагрузка в этот день в 2003 году составила 196%, 216%, 88% от средней интенсивности нагрузки для местной, междугородней и мобильной связи соответственно. В 2004 году данные значения составили 186%, 246%, 93%; в 2005 году соответственно 220%, 315%, 98%; в 2006 году – 245%, 343% и 99%. Особый интерес вызывает изменение объема и параметров нагрузки операторов мобильной связи, передаваемой по телефонным сетям и их соотношение с нагрузкой телефонных сообщений.

Данное исследование показывает основные тенденции изменения нагрузки на современных ТС, что еще раз подчеркивает то, что современные методы расчета объема сетевого оборудования не учитывают процессы, происходившие последние несколько десятилетий на ТС, не отражают перспективы конвергенции телефонных сетей с другими ТС, не учитывают реальную структуру потоков.

Заключение

В данной работе приведен лишь один пример анализа нагрузки различного вида на современной ТС. Данный пример наглядно показывает, что и количественно и качественно реальные параметры потоков сообщений на телекоммуникационных сетях существенно отличаются от модели, используемой при расчете количества оборудования ТС, применяемой проектными организациями в настоящее время, что доказывает важность и актуальность подобных исследований, которые могут решить ряд задач повышения качества предоставления услуг пользователям ТС. К таким задачам можно отнести:

- анализ влияния нетелефонной нагрузки на качество работы телефонной сети;
- анализ влияния нагрузки, создаваемой операторами мобильной связи, на качество работы телефонной сети;
- анализ корректности используемых методов расчета соединительных устройств на современных ТС;
- анализ соответствия применяемой модели потока вызовов потокам, циркулирующим на реальных ТС в настоящее время.

Естественно, приведенный перечень задач неполный, однако он позволяет делать вывод о необходимости и своевременности подобных исследований. Исследования параметров потоков вызовов и нагрузки на реальных ТС выполняются уже в течение длительного времени. Объем используемой выборки требует применения новых методов анализа, основанных как на использовании современных систем управления базами данных, так и специально разработанных для данных исследований программных продуктов. Существующие методы измерения параметров сообщений на ТС и исследования полученных результатов необходимо корректировать в соответствии с новыми задачами и новыми возможностями, что и является одной из целей дальнейших исследований.

Благодарности

Настоящая работа была выполнена при поддержке интернационального проекта *ITHEA XXI* Института информационных теорий и их приложений *FOI ITHEA* и Ассоциации *ADUIS* Украина (Ассоциация разработчиков и пользователей интеллектуальных систем).

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Литература

1. Гайворонская Г.С. Исследование параметров объединенного потока вызовов // Труды УГАС "Информатика и связь". – 1997. – с. 222-226
2. Гайворонская Г.С. Исследование параметров объединения информационных потоков различных видов связи в единый интегральный поток вызовов // Труды III Международной НТК по электросвязи, телевизионному и звуковому вещанию. – Одесса, 1997. – с. 128–239
3. Гайворонская Г.С. Объединение потоков нагрузки на сетях связи с обходными направлениями // Труды УГАС «Информатика и связь». – 1998
4. Гайворонская Г.С. Оценка параметров объединенного потока вызовов при помощи имитационного моделирования // Труды УНИИРТ. – 1998. – №2 (14). – с. 93-98

Информация об авторе

Илья Ганницкий – Одесская государственная Академия холода; аспирант кафедры информационно-коммуникационных технологий; ул. Дворянская, 1/3, Одесса – 82, 65082, Украина; e-mail: immortalbrain@gmail.com

Главные области научных исследований: потоки вызовов, нагрузка в современных телекоммуникационных сетях

PLAGIARISM REVEALING IN THE SOURCE PROGRAM CODE ON THE BASIS OF ITS SYNTACTIC REPRESENTATION

Tetyana Shatovska, Iryna Kamenieva

Abstract: *This paper considers the existing approaches and models of data representation applicable for the problem task of plagiarism revealing in the source code. This paper present an approach based on the solution concerning syntactic comparison of source codes for the task of plagiarism revealing in students course code. The first phase of analysis, this calculation of quantitative characteristics of the source code, in other words, an attribute algorithm. This preliminary calculation allows us to obtain an overall picture of project files, which is subjected to analysis. Attribute algorithm allows us to discard obviously not like a pair of source files in a relatively short period of time, resulting in we will get the abbreviated set of couples who want to compare the structural algorithm. The next phase is he analysis of potentially similar files of a structural algorithm. In this paper, a measure of similarity of these files have been selected as the value of the greatest common subtree. This problem has much in common with the problem of isomorphism of graphs.*

Keywords: *Plagiarism, source code, isomorphic graph, detector of plagiarism.*

ACM Classification Keywords: *D.2.5 - Testing and Debugging*

Introduction

With rapid development of mediums and forms of communications on the basis of modern information technologies such problem, as plagiarism expands also. It widespread almost in every fields of human activity. This fact not only breaks the copyrights of the person, whose work has been used as plagiarism, but also completely destroys the essence of education itself. In the sphere of education this problematic has global scales.

Education is a process of the direct interaction of the student and the lecturer, where the student makes efforts to obtain knowledge, and the lecturer in his turn helps to master the obtained knowledge, and as result, to estimate them adequately. Student's work containing the plagiarism cannot be estimated adequately in the main for the obvious reasons. In many cases the lecturer is not able to recognize the plagiarism and accepts the work as fulfilled, estimating it with a high mark and does not assume, that this student has appropriated the results of another's work.

As is known, students of technical universities in IT sphere practice the program implementation of algorithms, information systems and other programs within the course. The main objective of the tasks performing is to give knowledge and to increase the comprehension in the specific area. Unfortunately, many students resort to plagiarism usage in their works. For example, one of the students has taken a source code written by other student and has claimed it as his own or has downloaded a source code by means of the Internet resource.

In the modern world the Internet is called-for and keep on gaining the popularity among the students and consequently the access to various resources increases.

Since for today the problem is rather actual, for its solution there are already various automated approaches and resources which are successfully applied to search the plagiarism in essays, notes to the yearly projects, diplomas and other text documents. The great number of tools in the Internet allows to verify the text for plagiarism, but thus the plagiarism revealing in a program code is not studied thoroughly. The plagiarism revealing in the source code has almost nothing common with plagiarism revealing in the text as programs among themselves can have the identical syntactic representation, but thus to have the various text. We reach

the conclusion, that for the solution of the set problem the perfect automated resource which will allow to speed up the plagiarism revealing process is necessary. Further in paper the matters concerning the automated revealing of plagiarism in the source code will be considered.

Plagiarism in the source code

Before starting to look into the various ways of plagiarism revealing it is necessary to determine the understanding of this term. To define the quality of plagiarism searching this term should be describe. The plagiarism searching task as well as definition itself has wide enough and not accurate sense, depending on a subject field it can be perceived in different ways. The term plagiarism has various related statements:

- Plagiarism – Kind of copyrights violation, consists in illegal usage under someone own name of another's work (scientific, literary, musical) or invention, the innovation proposal (completely or partially) without indicating the source of adoption [1]. Compulsion to the co-authorship is also considered as plagiarism [2].
- Plagiarism – literal adoption from another's literary work without indicating the source [3].
- Plagiarism (from Lat. plagio – I abduct) – kind of author or the inventor rights violation. Consists in illegal usage under someone own name of another's work (scientific, literary, musical) or inventions, the innovation proposal (completely or partially) without indicating the source of adoption. [4]
- Plagiarism – appropriation of fruits of another's creativity: publication of another's works under someone own name without indicating the source or usage without the reformative creative changes made by the borrower. [5]
- Plagiarism – deliberate appropriation of authorship to another's work of science, literature or art. It is not considered as plagiarism the adoption of a theme or a story of work or the scientific ideas composing its content, without adopting the form of their expression. [6]

The above mentioned definitions are accurate enough and close to the essence but, nevertheless, they are not considered the specificity of plagiarism in the program code, therefore to put them into practice as the basic definitions of plagiarism is difficult. To understand which features should be considered for plagiarism searching in the program code we will observe, what is the computer program, as well as source code.

- Program (from Gr. programma – the declaration, the instruction, the decree) – the ordered sequence of operations for the computer, realizing the algorithm of some task solution [4].
- Source code (also the source text) is a text of a computer program in any programming language. In the generalized sense – any input information for the compiler. The source code is either compiled into executed code by means of the compiling program, or executed directly according to the text by means of the interpretive program [7].

Usually for the analysis of programs the source code containing the data required for the analysis is used. The source code is always linked to the programming language, namely with its grammar, therefore it is the structured text which has limitations set by grammar of the programming language. For this reason two programs having absolutely different text idea could be identical in syntactic representation, and, therefore, could be plagiarism. Let us consider the following instance:

<pre>public class A { private int a = 0; public int method(int b) { return a * b; }}</pre>	<pre>public class B { private int c = 0; public int anotherMethod(int d) { return c * d;}}</pre>
---	---

In this instance the elementary way of plagiarism hiding in the source code has been shown. It is obvious, that the source texts of the presented extracts do not coincide, but thus have absolutely identical logic and semantics. More detail the methods of plagiarism masking will be considered in the following sections of this paper.

The plagiarism searching in the program code demands the special approaches independent on text representation of the program, and also are steady to various ways of plagiarism masking.

Further we will consider the various models and approaches for plagiarism revealing.

Plagiarism searching in the source code

From the previous sections it becomes clear, that plagiarism searching in the source code cannot be realized by means of simple line-by-line comparison or by finding of the greatest common line, the plagiarism searching in the program code is complicated by that frequently they try to hide it. Now the following ways of plagiarism masking [8] are known:

- Code comments changing,
- Code formatting changing,
- Identifiers renaming,
- Change of operands order which are independent on execution time,
- Original data types substitution for related,
- Expressions substitution for identical in syntax,
- Operations and variables adding which are not in use or excessive,
- Code extracting from methods and on the contrary
- Original code combining with plagiarism
- Translation of the source code into other programming language

There are two basic approaches for plagiarism searching in the source code. First is based on program significant attributes counting, and the second is on structural representation of the source code.

The attribute method [8,9] is applied to digital expression of program significant attributes using further comparison of the obtained results with other analyzed programs. This method is capable to detect plagiarism in the masked source code. The significant attributes sampling depends on the programming language in which the source code is written. Usually using such approach the following basic attributes are considered:

- quantity of classes,
- quantity of methods,
- quantity of variables,
- quantity of loops,
- quantity of methods invocations,
- quantity of conditions,

The specifying attributes which are dependent on the programming language (quantity of abstracts, imports, enumerations and other specific features of the programming language) can be optionally used. In this approach it is very important to choose the correct metrics for further calculations to obtain proper result. Let us consider the instance showing the attribute method usage.

<pre>public class A { public int getSum(int[] array) { int sum = 0; for (int i = 0; i < array.length; i++) { sum += array[i]; } return sum; }}</pre>	<pre>public class B { public int addition(int[] array) { int sum = 0; int i = 0; if (true) { while (i < array.length) { sum += array[i++]; } } return sum; }}</pre>
---	--

From the resulted above code it is obvious, that both classes considerably differ from each other in their text representation. Class A contains 1 method, 1 loop and 3 declarations of a variable, including arguments of

method and a variable declared in the loop. Let us consider class B as potential plagiarism of A class. Class containing plagiarism is masked by operand substitution for identical in syntax and also by adding of a superfluous operand which does not change logic of code execution and contains 1 method, 1 loop, 1 condition and 3 declarations of a variable including input arguments. As a result of such calculation it is possible to say, that programs are very similar and potentially are plagiarism.

The main principle of such approach operation has been shown in this instance. In practice it may be complicated by adding of weights for different metrics etc. Such approach potentially can give the correct result, but it is also ineffective enough, as two absolutely different programs which differs both text and syntactic representations can give similar result that is not acceptable to solve our problem.

Next and most prospective approach consists in comparison of programs taking into account their structure [8,9]. Such procedure is much more difficult, than simple detection of quantitative characteristics of the program. This is due to the complexity of the analysis of the source code structure. In such approach each characteristic of the program is analyzed not isolated, but in context of its parent element. All units of the program are linked hierarchically and are at various levels of the enclosure. The structural method exceeds the attribute one by quality, but can be worse by productivity as the analysis of the program structure could demand much more calculations, than during conventional counting of its quantitative characteristics.

The main principle of our method is the preliminary construction of the source code tree, namely Abstract Syntax Tree [10] including its further analysis. Having built the Abstract Syntax Tree we obtain the complete conception about the source code, its units, structure, interconnections that give the possibility to realize the algorithms which are dependent on these data. On the other hand such approach is not flexible enough for the reason, that each programming language requires the separate parser which builds the Abstract Syntax Tree. Each programming language can contain various units in the structure which are defined by grammar of language, that directly influences the structure of the program tree, correspondingly the analyzer will be linked to the certain type of the Abstract Syntax Tree. This approach adaptation to any of programming languages demands the considerable efforts. These deficiencies are the pay for the high-quality approach. The classical instance of this approach realization consists in comparison of subtrees of various programs structure.

To understand the efficiency of the structural approach let us consider the following instance:

<pre>public class A { private String concat(char separator, String... strings) { StringBuilder sb = new StringBuilder(); for (String str : strings) { sb.append(str).append(separator); } return sb.toString();}}</pre>	<pre>public class B { class Inner { private String add(String[] values, char sep) { StringBuilder sb = new StringBuilder(); int i = 0; while (i++ < values.length) { sb.append(values[i]); sb.append(sep); } return sb.toString();}}</pre>
---	---

In the code of the program resulted above the class A is the original. Proceeding from it, it is obvious, that class B. Inner is plagiarism of the class A with insignificant alterations which do not influence the logic of program execution. For plagiarism masking the code has been wrapped in the internal class, the loop was substituted for identical one and the order of the method parameters declaration was changed. It is obvious, that text comparison of these source codes will not give demanded result. The text state of the program has been changed and differs from the original.

So, we need to build the structural representation of each of programs, namely Abstract Syntax Tree for each source code. Such procedure will allow to analyze the program from the point of view of its structure, having discard the superfluous information about the names of identifiers etc. Fig. 1 and 2 show the visualization of the tree structure of programs (original and plagiarism).

Original:

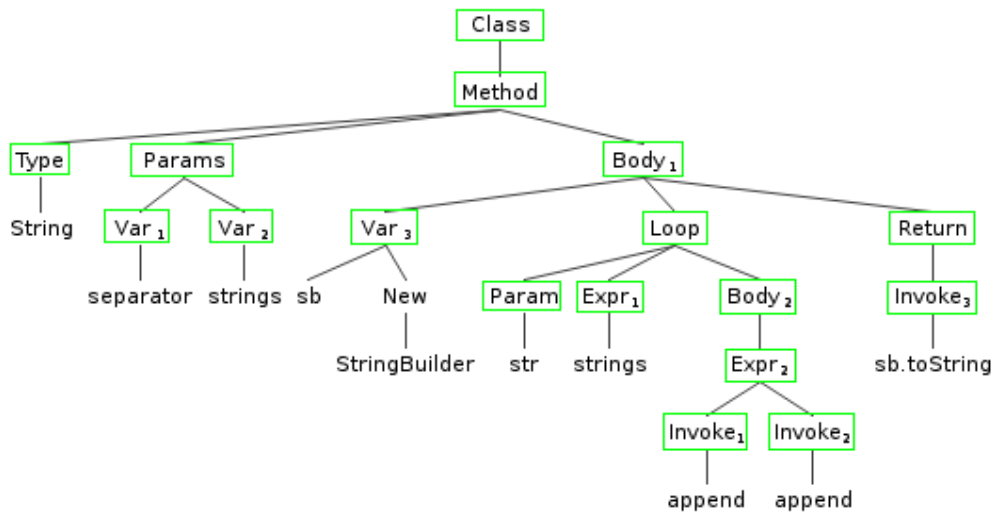


Figure 1 The original program tree structure visualization

Plagiarism:

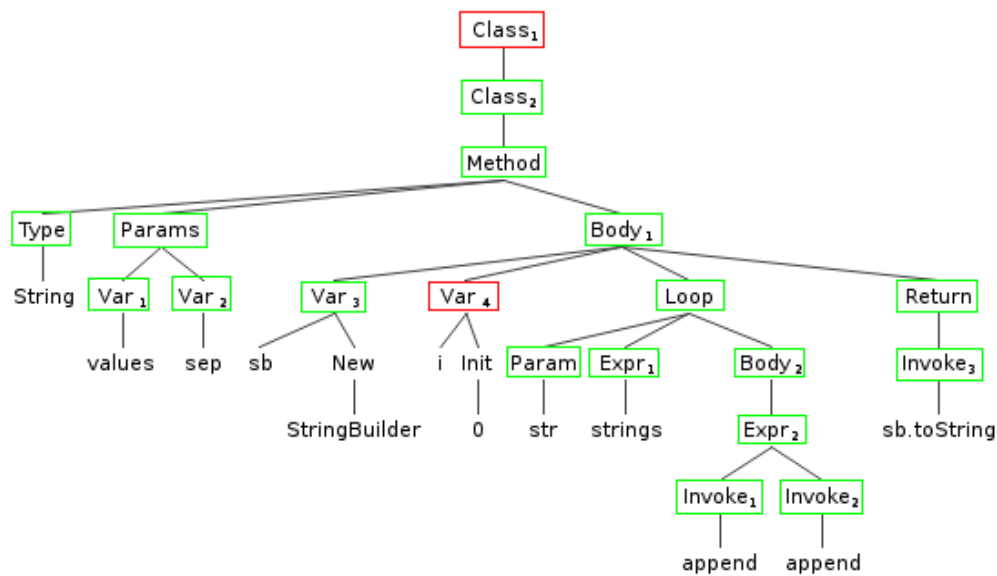


Figure 2 Plagiarism program tree structure visualization

Having analyzed both structures it is possible to conclude, that the structure of the original program is the subtree of the plagiarism program structure even taking into account such fact, that the code has been changed. Thus, we conclude, that program B is plagiarism. Programs can coincide not completely but only partially, small sections of one tree can coincide with small sections of the other tree, therefore, the common subtree can practically start with any unit. Proceeding from this fact, the complexity of any algorithm working under such principle increases, since it is necessary to search all units of the tree approximately so much times how many units in it. The algorithm complexity is cubic, that creates difficulties in practice usage because of low productivity. It also should be noted, that there are approaches for plagiarism detection in the source code which can work on the basis of neural networks, classifications and clusterizations.

Combined approach

To obtain the high efficiency of algorithm the combined approach for plagiarism revealing in the source code is used. As it has been described above, the attribute method has enough productivity, but in its turn can give absolutely incorrect result, the structural method allows to reveal the plagiarism precisely enough, but demands labour-intensive calculations. Proceeding from it, we conclude that there is combining of these two approaches.

The small programs consisting of one file have been presented above, but in fact almost any program consists of the numerous files great number of which forms the program.

Any program has both big and small files with the source code, which are required to be analyzed for plagiarism revealing. These can be both simple matters and the big files with the source code where the basic functionality, algorithms etc. are realized

For algorithm improving the following approach is offered. The first phase of the analysis is the counting of quantitative characteristics of the source code (attribute algorithm). The preliminary calculation allows to obtain the common information about project files. The first stage does not require the difficult calculations and it means, that we obtain total characteristic about the source code without considerable expenses for productivity. It is known, that two files with the source code can give equal or very close quantitative characteristics and thus to have absolutely different logic of performance, therefore, we cannot be assumed only on the results obtained during the first phase. The attribute algorithm allows to discard certainly not similar pairs of files with the source code for short enough period. Therefore we will obtain the reduced ensemble of pairs which should be compared by means of structural algorithm.

The second phase - the analysis of the potentially similar files by the structural algorithm. In this paper the measure of similarity of such files the value of the greatest common subtree had been chosen. This problem has much in common with the problem of graphs isomorphism. Isomorphism is the common concept which is used in various sections of mathematics. Generally it can be described as follows: let two ensembles with certain structure (groups, rings, linear spaces, etc.) is given. The bijective mapping between them is called isomorphism if it saves this structure. Such ensembles with structure is called the isomorphic ones. Isomorphism always sets the relation of equivalence on the class of such ensemble with structure. The objects between which there is the isomorphism, are in a sense "equally arranged", they are called isomorphic ones. As classical instance of isomorphic systems can be ensemble R of all real numbers with the operation of addition defined on it and ensemble R^+ of the positive real numbers with the operation of multiplication set on it. Representation $x \rightarrow \exp(x)$ in this case is isomorphism.

During plagiarism identification in isomorphic graphs the main aim becomes the identification of largest common substructure. The size value of the common substructure, in our case it means subtree, determines the level of similarity between analyzed programs, the more value the more programs are similar.

Architecture of the application

Let us consider the common architecture of the offered detector of plagiarism, the components of which the application and units with which it will work should consist. Let us start from the description of the operation principle of the plagiarism detector. On the entry the set of files with the source code is feed, then these files should be present as a tree structure to be analyzed. As it has been told earlier, it is offered to use the combined approach to analyze the source code for plagiarism revealing. Therefore, in the considered architecture there is provided two phases of the analysis. The first and the fastest phase is based on counting of significant ensembles and allows to discard certainly not similar pairs, and the second one analyses the structure of potentially similar files. The system kernel could be divided at least into three basic components.

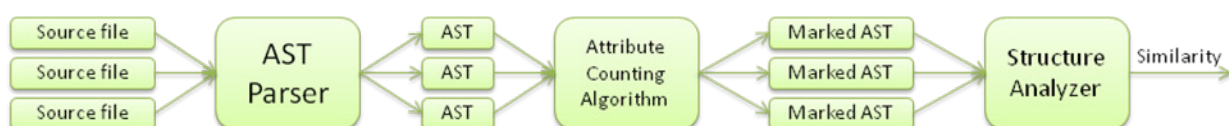


Figure 3 Main modules of the system

In the figure above you can see main modules of the system and the common sequence of performance. Let us consider the system components.

AST parser is required to construct Abstract Syntax Tree of the source code. This model of the data presentation is necessary for the analyzer of the programmed application. Parser is a front part of the application through which all input data pass. Then data are converted into required type, that is into AST. This part of the application is in any compiling program, compilation of any program begins from the construction of the source code tree.

Attribute Counting Algorithm is a model where the quantitative characteristics of the program are calculated by means of attribute algorithm on which entry the ensemble of the AST representations moves. As a result each tree is marked by the characteristics calculated for it and transferred to the next module.

Conclusions

On the basis of studies has been proposed and developed an hybrid approach to solve the plagiarism problem of the source code. The developed software tool has both scientific and practical value, hence the area of application. In terms of practices developed application can be used to search for plagiarism among a set of student projects with source code. In terms of scientific value, the result of the research, as well as evaluation of results can be used for further development of the developed approach.

Aknowledgement

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.itheta.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Bibliography

- [Синельников et al, 1994] С. Синельников, Т. Соломоник, М. Биржаков. Энциклопедия предпринимателя – Санкт-Петербург: ОБИС, 1994.
- [Сухарев, 2002] А.Я Сухарев. Большой юридический словарь – Москва: ИНФРА-М, 2002.
- [Блокгауза et al, 2001] Ф.А. Блокгауза, И.А. Ефрона. Энциклопедический Словарь – 86 т., С.-Петербург: 1890 – 1907// <http://slovari.yandex.ru/dict/brokminor/article/23/23998.html?text> , 2001.
- [Прохоров et al, 1978] Большая советская энциклопедия – 3-е издание: 30 томов, под ред. А.М. Прохоров и др. // <http://bse.sci-lib.com>, 1969 – 1978гг.
- [Асмус et al, 1939] В. Ф. Асмус, Д. Д. Благой, Б. М. Гранде и др. Литературная энциклопедия под ред. В.М. Фриче, А.В. Луначарский и др. // <http://slovari.yandex.ru/dict/litenc> ,1929-1939 гг.
- [<http://www.glossary.ru/>, 2005] Тематический толковый словарь // <http://www.glossary.ru/> , 2005.
- [Wikipedia] Wikipedia // http://ru.wikipedia.org/wiki/Исходный_код, см. 2010.
- [Hamilton et al, 2008] J. Hamilton, S. Danicic. Static Source Code Analysis Tools and their Application to the Detection of Plagiarism in Java Programs // <http://whoyouknow.co.uk/uni/msci/report.pdf>, 2008.
- [Goel et al, 2008] S. Goel, D. Rao et. al. Plagiarism and its Detection in Programming Languages // <http://stanford.edu/~drao/Resources/Plagiarism.pdf> , 2008.
- [Aho et al., 2007] Alfred V. Aho, Monica S. Lam, Ravi Sethi, Jeffrey D. Ullman. Compilers: Principles, Techniques, and Tools. (2nd Edition). Addison-Wesley, CA – 2007г.

Authors' Information



Tetyana Shatovska – Ass. Ph.D. Prof., Kharkiv National University of radioelectronics; Lenina av. 14, 61166 Kharkiv, Ukraine; e-mail: shatovska@gmail.com
Major Fields of Scientific Research: Data Mining, Text Classification, Business Intelligence, Artificial Intelligence



Iryna Kamenieva – Ph. D student., Kharkiv national university of radioelectronics; Lenina av. 14, 61166 Kharkiv, Ukraine; e-mail: iryna.kamenieva@gmail.com
Major Fields of Scientific Research: Data Mining, Text Classification, Business Intelligence, Artificial Intelligence

АЛГОРИТМ СИНХРОНИЗАЦИИ ОБЪЕКТОВ РАСПРЕДЕЛЕННОЙ ИМИТАЦИОННОЙ МОДЕЛИ В TRIAD.NET

Елена Замятина, Сергей Ермаков

Аннотация: В работе приводится краткий обзор классических алгоритмов синхронизации процессов в распределенных системах имитации, рассматривается распределенная система имитации Triad.Net и особенности реализации алгоритма синхронизации процессов в этой системе. При реализации алгоритма синхронизации авторы использовали знания о модели (знания пользователей о модели, знания, извлекаемые из структуры модели и знания, извлекаемые в процессе функционирования модели). Знания о модели представлены в виде правил. Алгоритм основан на классическом оптимистическом алгоритме. В работе приводятся результаты экспериментов, которые показали эффективность предложенного алгоритма синхронизации.

Keywords: имитационное моделирование; экспертные системы, распределенное имитационное моделирование, консервативный алгоритм синхронизации, оптимистический алгоритм синхронизации, база знаний.

ACM Classification Keywords: I.6 Simulation and Modelling – I.6.2 Simulation Languages; I.2 Artificial Intelligence – I.2.5 Programming Languages and Software – Expert system tools and techniques.

Введение

В настоящее время актуален переход к созданию распределенных систем имитационного моделирования (СИМ). Как известно, имитационное моделирование является часто используемым, а иногда и единственным методом исследования сложных систем. Сложность задач, решаемых методом имитационного моделирования, явилось причиной, по которой наряду с последовательными системами имитационного моделирования появились распределенные/параллельные системы имитации [Fujimoto, 2003; Meyer, 1998; Окольнішников, 2005]. Разработка распределенных систем имитации сопровождается необходимостью создания специального программного обеспечения, которое позволило бы использовать ресурсы гетерогенных или гомогенных многопроцессорных или мультимедийных вычислительных систем (ВС). Использование ресурсов нескольких вычислительных узлов во время имитационного эксперимента сокращает время его проведения, однако алгоритм управления объектами, распределенными по вычислительным узлам (в дальнейшем будем называть его алгоритмом синхронизации), является гораздо более сложным, чем алгоритм продвижения времени в последовательном моделировании. Время выигрыша от использования ресурсов нескольких вычислительных узлов может быть увеличено за счет эффективного алгоритма синхронизации. В докладе авторов предпринята попытка ускорить время выполнения алгоритмов синхронизации за счет его адаптации к конкретной имитационной модели, за счет знаний о ее поведении, которые используются при проведении распределенного имитационного эксперимента.

Алгоритмы синхронизации

Традиционно используются два подхода к реализации алгоритма синхронизации объектов моделирования, распределенных по разным вычислительным узлам: консервативный и оптимистический. Основной целью этих алгоритмов является обеспечение выполнения каждым логическим процессом событий в порядке не убывания их временных меток, чтобы сохранить причинно-следственные связи.

Принципиальная задача консервативного алгоритма – определить время, когда обработка очередного события из списка необработанных событий является «безопасным». Событие является безопасным, если можно гарантировать, что процесс в дальнейшем не получит от других процессов сообщение с меньшей временной меткой. Консервативный подход не позволяет обрабатывать событие до тех пор, пока не убедится в его безопасности. Подробное описание консервативных алгоритмов можно найти у Р.Фуджимото, Чанди, Мизры и др. [Bryant, 1977; Chandy, 1978].

В отличие от консервативных алгоритмов, не допускающих нарушения ограничения локальной каузальности, оптимистические методы не накладывают такого ограничения. Они обеспечивают восстановление порядка при его нарушении. Самый известный алгоритм - Time Warp, который был разработан Джефферсоном [Jefferson, 1990]. Когда логический процесс получает событие, имеющее временную отметку меньшую, чем уже обработанные события, он выполняет откат и обрабатывает эти события повторно в хронологическом порядке. Откатываясь назад, процесс восстанавливает состояние, которое было до обработки событий (все состояния системы сохраняются) и отказывается от сообщений, отправленных в результате обработки событий, которые необходимо отменить в результате отката. Для отката от этих сообщений разработан механизм антисообщений.

Консервативные и оптимистические алгоритмы представляют собой две крайности. Для некоторых специфических моделей консервативные алгоритмы показывают производительность, превосходящую оптимистические в несколько раз, обратное утверждение, вообще говоря, тоже верно. Однако, проанализировав все перечисленные выше алгоритмы синхронизации, можно четко проследить тенденцию возрастания эффективности алгоритма при увеличении количества информации, которое известно о модели (lookahead, lookback, временная метка следующего события и т.п.). Эту информацию используют как консервативные, так и оптимистические алгоритмы: консервативные, таким образом, позволяют увеличить горизонт безопасных событий, а оптимистические наоборот, сдержать чрезмерное продвижение модельного времени. Отсюда можно сделать вывод, что лучший алгоритм должен сочетать в себе оба подхода. Сформулируем требования, которым должна удовлетворять информация:

- Информация помогает максимально повысить эффективность алгоритма имитации.
- Информацию может вводить пользователь (создатель модели, эксперт).
- Информацию можно получать от самой системы моделирования в процессе её выполнения.
- Информацию легко хранить и обрабатывать.
- Информация адекватно представляет знания о модели.

В настоящей реализации информация о модели представлена в виде продукционных правил.

Рассмотрим алгоритм, использующий такую информацию во время имитационного прогона и созданный на базе оптимистического алгоритма. С точки зрения моделирования, главным является соблюдение правильного порядка выполнения модели во времени. Следовательно, получаем два основных типа связи, который сами по себе связаны друг с другом:

- Каузальная связь - связь последовательности выполнения событий.
- Дистанция по времени (временное расстояние) между событиями.

Данные двух типов связей применимы к объектам, имеющим временные метки, то есть к событиям и сообщениям.

Пользовательская информация

Как правило, пользователь системы моделирования имеет некоторое приблизительное неточное представление о модели, которое он может описать в виде правила, например:

«Прежде, чем получить товар, его нужно оплатить» - получаем представление этого знания в виде правила: «IF Оплатить Товар THEN Получить Товар». Как видно из примера, информация, задаваемая пользователем, является достаточно простой, однако для алгоритма синхронизации она позволит сократить количество откатов, и, следовательно, уменьшить время выполнения модели.

Очевидно, что знания эксперта являются нечеткими и ненадежными (например, в данном конкретном случае, товар может быть украден, или наоборот, после оплаты оказалось, что товар некачественный и подлежит возврату), поэтому каждому правилу приписывается коэффициент доверия CF.

Казуальная зависимость между событиями в общем виде представляется в виде

$$\text{IF } E_1 \text{ AND } E_2 \text{ AND } E_3 \text{ AND } \dots \text{ AND } E_n \text{ THEN } E_k \text{ CF } [0..100].$$

Здесь E_k зависит от $E_1, E_2, E_3, \dots, E_n$. CF - коэффициент доверия. 0 – нет доверия, 100 – максимальное доверие.

Информация, генерируемая самой системой

Пользовательская (экспертная) информация является самой важной, так как эти знания мы получаем от эксперта в своей предметной области. Однако необходимость детального описания поведения системы пользователем затруднительна для самого пользователя (извлечение знаний – это трудоемкий процесс). К тому же при имеющемся всеохватывающем описании системы теряется главный смысл проведения моделирования – зачем его проводить, если мы и заранее знаем поведение системы. Исходя из этого, система должна уметь сама обеспечивать себя необходимой информацией. Для упрощения описания информации и для интегрирования двух типов знаний, системная информация, так же как и пользовательская будет представлена в виде правил. Системные правила формируются на различных стадиях:

- *На стадии трансляции:*
рутина представляет собой пользовательские процедуры обработки событий и сообщений, поэтому она по определению содержит информацию о каузальной связности событий и сообщений.
- *На стадии выполнения:*
во время выполнения модели последовательность обработки событий и сообщений анализируется, и на основе этой информации генерируются правила.

В общем случае, при увеличении количества различных событий и сообщений в системе, количество правил может расти экспоненциально. Поэтому генерировать все возможные зависимости очень трудоемко для системы. Более того, большинство из сгенерированных правил могут ни разу не сработать (например, если правило описывает отношения объектов в рамках одного и того же локального процесса).

Поэтому, исходя из потребностей алгоритма моделирования в информации о связи событий, будем придерживаться следующих утверждений:

- Генерация правил происходит только для объектов системы находящихся в разных локальных процессах.
- Наиболее значимыми правилами являются те правила, которые помогают предотвратить временной парадокс (они генерируются из тех событий, которые в прошлом вызвали временной парадокс).

Сбор, анализ и распространение информации

Опишем общую схему реализации сбора информации. Сам по себе логический процесс не должен тратить свои вычислительные ресурсы на обработку информации. Он должен использовать ее уже в готовом виде. Поэтому предлагается создание специального агента, который выполняет следующие функции:

- Сбор информации.
- Удаление избыточной информации.
- Анализ информации на непротиворечивость.
- Преобразование информации в удобный для использования вид (сворачивание мелких правил в одно более общее, сортировка правил в соответствии с их точностью (CF) и релевантностью).
- Обмен информацией между локальными процессами.

Кроме агентов, расположенных на локальных вычислительных узлах, вся информация агрегируется в центральную базу знаний, расположенную на отдельно выделенном сервере. Главный управляющий процесс выполняется на этом же сервере. В качестве агентов на локальных вычислительных узлах используют информационные процедуры. Речь о них пойдет несколько позже. Информационные процедуры – это средства сбора информации об имитационной модели в Triad.

Использование информации во время имитационного прогона

Ключевой момент алгоритма – использование полученной информации о модели для синхронизации объектов имитационной модели во время прогона. Поскольку консервативный алгоритм синхронизации не допускает временных парадоксов, а применение правил не дает уверенности в данном случае неприемлемо. Алгоритм в системе Triad.Net будет строиться на основе обычного оптимистического алгоритма с посылкой анτισообщений и откатами (алгоритм Time Warp). Главная задача алгоритма – минимизировать количество откатов в системе, которые тормозят работу системы, при максимальном сохранении параллелизма в модели. Чтобы откаты не происходили, необходимо, чтобы обрабатывались только безопасные события.

Рабочая память системы содержит факты – события и сообщения, которые были обработаны в течение некоторого промежутка времени ΔT_{act} это время, в течение которого данные объекты считаются актуальными. Каждый раз при увеличении времени локального процесса, рабочая память актуализируется.

Определение безопасности события происходит по следующему алгоритму:

- Выполняется поиск данного события в заключениях всех правил в базе знаний системы.
- Если таких правил не обнаружено, то происходит выполнение по обычному сценарию.

Если же такие правила обнаружены, то выполняется логический вывод по продукционной базе знаний. В результате возможно 2 результата: событие В выведено с некоторой вероятностью CF или событие В не выведено. При вычислении доверия к заключению доверия к сработавшим правилам комбинируются.

Важно заметить, что изначально база знаний может быть не заполнена до такой степени, чтобы обеспечить эффективное выполнение алгоритма: если событие В не было получено в результате логического вывода, то это не значит, что его не надо обрабатывать в данный момент. Это может значить, что еще не было добавлено правило, в результате которого можно получить событие В в текущем состоянии системы. Поэтому необходимо установить временную границу начального обучения алгоритма, когда его база знаний (БЗ) наполняется. Если система находится в стадии заполнения БЗ, то событие В так или иначе будет обработано, в противном случае, не будет.

Система решает, следует ли обрабатывать данное событие, исходя из доверия к заключению. Возможны две стратегии решения обрабатывать событие или нет:

- *Пороговая*: если CF больше заданного порога (для каждого объекта может быть свой порог: статический или динамически вычисляемым (например, с помощью генетического алгоритма)).
- *Вероятностная*: чем больше CF, тем больше вероятность, что событие будет обработано. В реализации этого подхода используют генератор случайных чисел.

В случае, если сработавшее правило привело к откату системы, его CF снижается. В противном случае, CF повышается.

В посылке и заключении правила можно указывать событие конкретного объекта (в случае, если в модели есть несколько одинаковых объектов): IF O1.E1 THEN O2.E2. Так же в посылке правила можно задать условие на состояние объекта: IF O1.Attribute1 = value1 AND O1.E1 THEN O2.E2. Это необходимо, в случае, если последовательность обработки событий зависит от внутреннего состояния объекта.

Таким образом, система позволяет максимально учитывать зависимости между объектами на стадии выполнения.

Описанный выше алгоритм, основанный на знаниях, реализован в системе Triad.Net[7]. Несколько слов о системе моделирования Triad.Net.

Распределенная система имитации Triad.Net

На кафедре математического обеспечения ВС разрабатывается распределенная версия системы автоматизированного проектирования и моделирования Triad.Net. Распределённая версия реализуется с использованием технологии .Net. Использование технологии .Net дают возможность реализовать гибкую компонентно-ориентированную систему.

Triad.Net является распределенной версией системы моделирования Triad [Миков, 2009; Миков, 2008]. Особенности системы моделирования:

- Входной язык описания моделей содержит переменные типа «модель». Над моделями определены операции. Операции определены как для моделей в целом, так и для каждого слоя (в Triad модель представлена слоем структур, рутин и сообщений.). Имитационная модель может быть описана средствами языка Triad или построена в результате исполнения некоторого алгоритма преобразования модели. При выполнении операций над моделью (или операций внутри каждого из слоев) перетрансляции модели не требуется.
- Модель является иерархической, т.е. каждая вершина в слое структур может быть расшифрована подструктурой.
- Подсистема анализа модели должна обеспечить получение информации по заранее сформулированному запросу, а не ограничивать пользователя строго регламентированным набором собираемых данных. Такой подход к сбору информации позволяет избежать избыточности собранной информации или того, что она окажется недостаточной.
- Алгоритм имитации должен быть эффективным и масштабируемым, т.е. объекты должны быть таким образом распределены по компьютерам, чтобы снизить потери, связанные с передачей сообщений по каналам связи от компьютера (процессора) к компьютеру (процессору) и учитывать загрузку компьютеров (процессоров) (В СИМ Triad разрабатывается подсистема управляемой динамической балансировки).
- Система должна быть объектно-ориентированной (должно поддерживаться наследование, переиспользование кода).

Описание имитационной модели

Описание имитационной модели в Triad состоит из трех слоёв: слоя структур (**STR**), слоя рутин (**ROUT**) и слоя сообщений (**MES**). Таким образом, модель в системе Triad можно определить как $M = \{STR, ROUT, MES\}$.

Слой структур представляет собой совокупность объектов, взаимодействующих друг с другом посредством посылки сообщений. Каждый объект имеет полюса (входные Pin и выходные Pout), которые служат соответственно для приёма и передачи сообщений. Слой структур можно представить графом. В качестве вершин графа следует рассматривать отдельные объекты. Дуги графа определяют связи между объектами.

Объекты действуют по определённому алгоритму поведения, который описывают с помощью рутины (rout). Рутинa представляет собой последовательность событий e_i , планирующих друг друга ($e_i \in E, i = \overline{1, n}$, E -множество событий, множество событий рутины является частично упорядоченным в модельном времени). Выполнение события сопровождается изменением состояния q_k объекта.

Состояние объекта определяется значениями переменных var_j рутины ($var_j \in Var, j = \overline{1, m}, Var$ – множество переменных рутины). Таким образом, система имитации является событийно-ориентированной. Рутинa так же, как и объект, имеет входные (*Prin* и выходные (*Prout*) полюса. Входные полюса служат соответственно для приёма сообщений, выходные полюса – для их передачи. В множестве событий рутины выделено входное событие *in*. Все входные полюса рутины обрабатываются входным событием. Обработка выходных полюсов осуществляется остальными событиями рутины. Для передачи сообщения служит специальный оператор *out* (*out* <сообщение> *through* <имя полюса>). Совокупность рутин определяет слой рутин *ROUT*.

Слой сообщений (MES) предназначен для описания сообщений сложной структуры. Система *Triad* реализована таким образом, что пользователь может описать только один слой. Так, если возникает необходимость в исследовании структурных особенностей модели, то можно описать в модели только слой структур. Алгоритмом имитации назовём объекты, функционирующие по определённым сценариям, и синхронизирующий их алгоритм.

Итак, выше были приведены сведения об имитационной модели в *Triad.Net* для выполнения ее на одном вычислительном узле. В распределенном варианте (выполнение на нескольких вычислительных узлах) имитационная модель представляет собой совокупность объектов, распределенных по разным вычислительным узлам. На одном узле могут располагаться один или несколько объектов. Во время имитационного выполнения объекты обмениваются информацией друг с другом, посылая сообщения. размещение объектов по вычислительным узлам происходит во время этапа статической балансировки. Критерием размещения объектов по узлам являются структурные особенности имитационной модели: сильносвязанные объекты целесообразно разместить на одном узле.

Информационные процедуры

Для сбора, обработки и анализа имитационных моделей в системе *Triad.Net* существуют специальные объекты – информационные процедуры и условия моделирования. Информационные процедуры и условия моделирования реализуют алгоритм исследования.

Информационные процедуры ведут наблюдение только за теми элементами модели (событиями, переменными, входными и выходными полюсами), которые указаны пользователем. Если в какой-нибудь момент времени имитационного эксперимента пользователь решит, что следует установить наблюдение за другими элементами или выполнять иную обработку собираемой информации, он может сделать соответствующие указания, подключив к модели другой набор информационных процедур. Информационные процедуры являются единственным средством системы для одновременного доступа к элементам модели, принадлежащим разным объектам. Именно с помощью информационных процедур пользователь может осуществить взаимодействие с объектами модели во время имитации.

Условия моделирования анализируют результат работы информационных процедур и определяют, выполнены ли условия завершения моделирования.

Система имитации *Triad.Net* располагает языковыми средствами для описания алгоритмов работы информационных процедур. В описание информационных процедур входят: описание начальной части (выполняется до начала имитационного эксперимента), описание заключительной части (эта часть выполняется по окончании эксперимента и служит для окончательной обработки интегральных характеристик, например, подсчет среднего значения), описания основной части информационных процедур, настроечных параметров и параметров интерфейса. При изменении значения переменной, за которой ведётся наблюдение, при выполнении события, указанного пользователем, или после прихода (передачи) сообщения на входной полюс происходит подключение информационной процедуры (ее основной части) к конкретному элементу модели (посредством параметров интерфейса) и данные обрабатываются по заданному в информационной процедуре алгоритму.

Таким образом, подсистема анализа модели обеспечивает получение информации по заранее сформулированному запросу, а не ограничивает пользователя строго регламентированным набором собираемых данных. Такой подход к сбору информации позволяет избежать избыточности собранной информации или того, что она окажется недостаточной.

Информационные процедуры и условия моделирования используют и для сбора информации о поведении модели, о ее характеристиках для организации алгоритма синхронизации.

Вычислительный эксперимент

Модель может быть представлена графически или на встроенном языке TRIAD [Миков, 1995]. Далее выполняется трансляция описания модели в объектный код. По окончании трансляции выполняется этап статической балансировки: размещение объектов имитационной модели по узлам вычислительной системы (ВС), ну а потом - запуск модели на выполнение.

Для проведения вычислительного эксперимента были выбраны несколько моделей, которые тестировались с применением различных алгоритмов на разных конфигурациях ВС с целью выявления наиболее оптимального алгоритма. Основной показатель оптимального алгоритма – время его выполнения, а для оптимистического алгоритма – это еще и количество откатов. Стоит отметить, что исследования выполнялись для достаточно простых (по вычислительной сложности) моделей. Время выполнения всей модели – максимальное из времен выполнения каждого локального процесса. Ниже приводится описание модели и результаты вычислительного эксперимента.

Модель вычислителя ILLIAC

Описание модели:

Модель представляет собой сильно упрощенную схему распределенных вычислений. Представлена мультипроцессорная вычислительная система с 64 вычислительными узлами – процессорами, где каждый процессор связан с четырьмя соседними. Процессоры выполняют вычисления и обмениваются данными между собой. Схема вычислительной системы изображена на рис. 1.

Анализ модели:

На каждый вычислительный узел попадает несколько процессоров. В самом лучшем варианте число внешних связей в системе будет $\frac{3}{4} * N$, где N – количество процессоров. Поскольку в системе сообщения передаются асинхронно, и количество связей велико, следовательно, количество возможных временных парадоксов и, как следствие, откатов (для оптимистического алгоритма) будет достаточно большим.

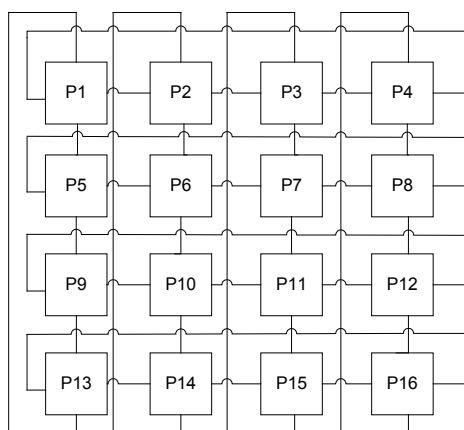


Рис. 1. Схема взаимодействия узлов в вычислительной системе ILLIAC

Эксперимент с моделью:

Проведем вычислительный эксперимент с моделью, для уменьшения объема результатов, будем полагать, что модельное время = 5000 ед.

Результаты эксперимента с моделью "ILLIAC"

Таблица 1.

Количество узлов	Время выполнения		
	оптимистический	консервативный	TriadRule
2	370,5	330,5	280,73
3	290,1	260,07	230,75
4	185	177,6	162,8
6	170,6	146,3	133,275
8	120	115	103,75
12	96,37	87,57	78,77
16	85	76,7	68,3
20	84,3	76,3	68,3
24	85,7	55,4	47,8
28	90	53,1	44,1
32	91,2	52,1	42,3
48	100	49,7	36,7
64	97,3	46	31,3

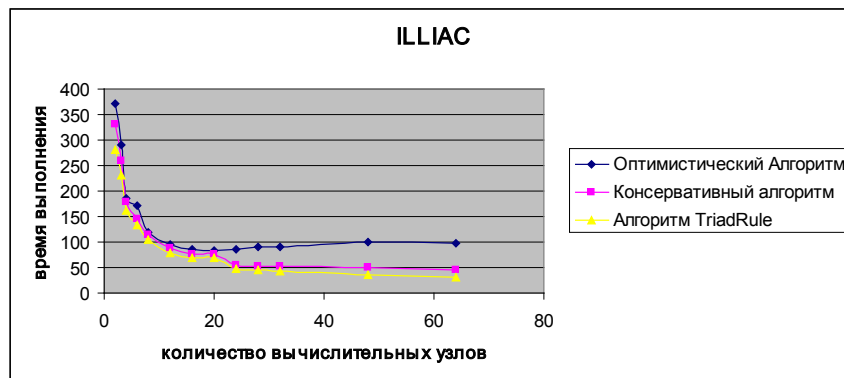


Рис. 1. Диаграмма Зависимости времени выполнения от количества узлов для модели ILLIAC

График показывает, что время выполнения всех алгоритмов за исключением оптимистического уменьшается с увеличением количества узлов. При увеличении количества узлов, количество внешних связей растет, а это приводит к увеличению количества откатов, следовательно, время выполнения оптимистического алгоритма, начиная с некоторой точки (в данном эксперименте это 23 вычислительных узла), растет. Консервативный алгоритм постепенно сокращает время выполнения, однако в силу ограниченности параллелизма, сокращение времени его выполнения происходит медленно. Новый алгоритм показывает самые лучшие результаты, поскольку позволяет сохранить параллелизм, уменьшив количество откатов в системе.

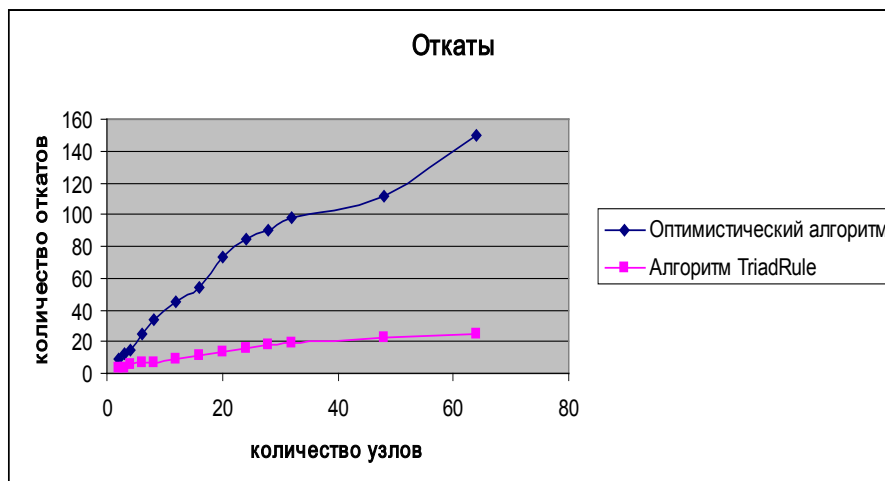


Рис. 2. Диаграмма зависимости количества откатов от количества узлов для модели ILLIACK.

Количество откатов в оптимистическом алгоритме возрастает прямо пропорционально увеличению количества вычислительных узлов, а, значит, и связей между локальными процессами. Поэтому, не смотря на ускорение вычислений, общее время выполнения увеличивается. Разработанный алгоритм позволяет свести возрастание количества откатов к минимуму, по графику видно, рост количества откатов происходит медленно.

Проанализировав полученные результаты можно заметить следующее: консервативный алгоритм сдерживает выполнение, уменьшая скорость выполнения, оптимистический алгоритм наоборот, слишком забегает вперед, что приводит к откатам, которые сильно тормозят продвижение времени. Алгоритм TriadRule – это оптимистический алгоритм, однако для сдерживания чрезмерного забега логического процесса вперед он использует правила. Эксперименты показали, что сдерживание получается минимальным. Однако для моделей, которые по своей сути являются параллельными, а именно, временные парадоксы в которых возникают редко или вообще не возникают, данный алгоритм оказывается хуже оптимистического. Рассмотрим этапы выполнения алгоритма TriadRule и их характеристику:

Этапы выполнения алгоритма синхронизации, основанного на знаниях

Таблица 2.

Этап	Скорость выполнения	Количество правил	Происходящие процессы
Начальный	Как у оптимистического	Правил нет, или очень мало	Накопление правил
Средний	Скорость приближается к консервативному	Правила есть, однако они срабатывают редко, что приводит к замедлению выполнения	Накопление правил и коррекция коэффициентов доверия
Последний	Скорость значительно выше и консервативного и оптимистического алгоритма	База знаний правил почти полностью сформирована,	Происходит лишь слабое пополнение БЗ и небольшая корректировка коэффициентов

Таким образом, алгоритм начинает приносить выгоду только на последнем этапе. Для каждой модели данный этап наступает в разное время. Главная задача повышения эффективности вычислений в том, чтобы последний этап наступил как можно скорее. Для того, чтобы приблизить этот этап, необходимо изначально наполнять базу знаний пользовательскими правилами.

Заключение

В работе рассматривается проблема разработки алгоритма синхронизации объектов распределенной имитационной модели, основанного на знаниях. Авторы предлагают модернизировать классический оптимистичекый алгоритм с откатами Time Warp. Выполнение классического оптимистического алгоритма корректируется с помощью продукционных правил, которые построены на знаниях пользователя о модели. Правила синхронизации включают также знания о модели, извлеченные во время функционирования модели.

В работе подробно описывается организация подсистемы синхронизации объектов имитационной модели, проблемы сбора информации о модели во время имитационного прогона (с использованием механизма информационных процедур), описание вычислительного эксперимента с алгоритмом синхронизации TriadRule и приводятся результаты этого эксперимента. Во время эксперимента определяется время имитационного прогона. Результаты эксперимента подтверждают, что разработанный алгоритм TriadRule позволяет частично сократить временные затраты на моделирование.

Сбор информации о модели выполняется с помощью информационных процедур. Для реализации алгоритма TriadRule были использованы встроенные информационные процедуры. Однако пользователь может внести в процедуру сбора информации коррективы, воспользовавшись языком Triad и лингвистическими средствами описания информационных процедур. Таким образом, алгоритму придается дополнительная гибкость.

Библиографический список

- [Fujimoto, 2003] Fujimoto R.M.. Distributed Simulation Systems. In Proceedings of the 2003 Winter Simulation Conference S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds. The 2003 Winter Simulation Conference 7-10 December 2003. The Fairmont New Orleans, New Orleans, LA, pp. 124-134
- [Meyer, 1998] Meyer R.A., Bagrodia R. Parsec User Manual. Release 1.1., UCLA Parallel Computing Laboratory, 1998
Доступно на сайте: pcl.cs.ucla.edu/projects/parsec
- [Окольнишников, 2005] Окольнишников В.В. Представление времени в имитационном моделировании. Вычислительные технологии. Т. 10, №5, Сибирское отделение РАН, 2005, стр. 57-77
- [Bryant, 1977] Bryant R.E., Simulation of Packet Communications Architecture Computer Systems. MIT-LCSTR-188. 1977.
- [Chandy, 1978] Chandy K.M., Misra J. Distributed simulation: a case study in design and verification of distributed programs IEEE Transactions on Software Engineering. 1978. Vol. SE-5(5).P. 440-452.
- [Jefferson, 1990] Jefferson D.R., Virtual time II: storage management in distributed simulation Proc. of the Ninth Annual ACM Symposium on Principles of Distributed Computing. 1990. P. 75-89.
- [Миков, 2008] Миков А.И., Замятина Е.Б. Технология имитационного моделирования больших систем // Труды Всероссийской научной конференции «Научный сервис в сети Интернет» – М.: Изд-во МГУ, 2008. С.199-204.
- [Миков, 2009] Миков А.И., Замятина Е.Б., Козлов А.А. Оптимизация параллельных вычислений с применением мультиагентной балансировки // Труды международной научной конференции «Параллельные Вычислительные Технологии». Нижний Новгород – Челябинск, Изд. ЮУрГУ, 2009. С. 599-604.
- [Mikov, 1995] Mikov A.I. Simulation and Design of Hardware and Software with Triad// Proc.2nd Intl.Conf. on Electronic Hardware Description Languages, Las Vegas, USA, 1995. pp. 15-20.

Сведения об авторах

Елена Замятина – Пермский государственный университет, доцент кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, 614017, ул. Тургенева, 33-40; e-mail: e_zamyatina@mail.ru.

Сергей Ермаков – Пермский государственный университет, аспирант кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, ул. Стахановская 7-52; e-mail: choufler@gmail.com.

ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА ПОДДЕРЖКИ УПРАВЛЕНИЯ РИСКАМИ В РЕГИОНАЛЬНЫХ ПРОИЗВОДСТВЕННЫХ КОМПЛЕКСАХ

Галина Кутергина, Людмила Лядова, Наталья Фролова

Abstract: В работе исследуется современное состояние проблемы управления рисками. Анализируются типичные недостатки существующих систем управления рисками в коммерческих организациях и причины, не позволяющие реализовать эффективные системы управления рисками в крупных производственных комплексах. Рассматриваются проблемы, связанные с отсутствием полноценной методологии управления рисками, которая базировалась бы на использовании современных информационных технологий, адекватных потребностям современного информационного общества по управлению глобальными и региональными социально-экономическими процессами, системными рисками. Описывается подход к созданию информационно-аналитических систем поддержки риск-ориентированного управления. В состав системы предлагается включить средства интеллектуального поиска и анализа документов, предназначенные для автоматического поиска, реферирования и каталогизации информационных ресурсов, содержащих информацию о данной области. Эта подсистема основывается на использовании онтологических моделей предметной области: поиск информации, реферирование документов и их классификация и каталогизация осуществляются в соответствии с моделями, разрабатываемыми экспертами с использованием предметно-ориентированных языков, позволяющих пользователям работать в привычных терминах предметной области. Эксперты в области управления рисками на основе полученной информации получают возможность создавать и оптимизировать модели управления рисками. Одна из наиболее сложных проблем при создании моделей – выделение ключевых показателей мониторинга. Для снижения трудоемкости работы экспертов при разработке системы показателей предлагается использовать средства анализа данных, основанные на теории исчисления фактов, позволяющие выявлять существующие скрытые зависимости в массивах данных, представляющих динамику производственных комплексов (если эти данные доступны). Анализ найденных зависимостей позволяет выделить или уточнить перечень показателей, влияние которых на деятельность производственных комплексов максимально. Созданные типовые модели предлагается сохранять в репозитории моделей, открытом для использования разработчиками систем управления рисками. На основе найденных в репозитории типовых моделей, относящихся к выбранным отраслям и видам деятельности, типам и видам предприятий, специалисты в области риск-ориентированного управления могут создать и оптимизировать системы управления рисками в своих производственных комплексах, используя накопленные экспертные знания и средства исследования моделей для анализа и оценки систем управления рисками, а также выработки рекомендаций по их оптимизации.

Keywords: риск-ориентированное управление; мониторинг рисков; онтология; интеллектуальный поиск; Text Mining; системы поддержки принятия решений; DSS; аналитическая обработка данных; исчисление фактов, Business Intelligence; OLAP.

ACM Classification Keywords: H. Information Systems. H.3 Information storage and retrieval: H.3.5 Online Information Services – Web-based services; H.3.6 Library Automation – Large text archives; H.4 Information Systems Applications: H.4.2 Types of Systems – Decision support (e.g., MIS).

Conference topic: Business Informatics.

Введение

Риск сопровождает деятельность любой компании – воздействие негативных факторов приводит к тому, что ни один бизнес-процесс не может быть осуществлен с заведомо гарантированным результатом. Таким образом, условием развития всякой компании, производственного комплекса является необходимость управления рисками. Развитие информационных технологий (ИТ) ускоряет процесс глобализации, постепенной утраты исключительной роли государства в мировом экономическом пространстве. Полноценными участниками мировой конкуренции становятся отдельные регионы, транснациональные корпорации, крупные региональные производственные комплексы (РПК), кластеры. В связи с этим повышается роль разработки и внедрения новых технологий управления развитием регионов и региональных производственных комплексов, использования в процессе управления ими передового мирового опыта, научно-технических достижений. В качестве приоритетных задач обеспечения экономической безопасности России выделены развитие индустрии информационных и телекоммуникационных технологий, разработка концептуальных и программных документов, а также создание комплексной системы контроля над рисками [18].

Попытки менеджмента отдельных успешных российских промышленных компаний разработать механизмы и инструменты управления рисками пока не смогли обеспечить эффективного, системного управления ими. На наш взгляд, это происходит, прежде всего, из-за *отсутствия полноценной методологии управления рисками, которая базировалась бы на использовании современных информационных технологий, адекватных потребностям современного информационного общества по управлению глобальными и региональными социально-экономическими процессами, системными рисками*. Технологии управления рисками не вполне соответствуют масштабам и характеру современных угроз устойчивого развития, требованиям инновационного типа развития производственных комплексов и регионов, тенденций международного разделения и кооперации общественного труда как важнейших факторов развития РПК. В производственных комплексах (например, авиационно-промышленный, нефтегазовый комплексы России) подобная практика находится лишь на самом начальном этапе осмысления необходимости формирования системы управления рисками.

В представленной работе анализируются типичные недостатки существующих систем управления рисками в коммерческих организациях и причины, не позволяющие реализовать эффективные системы управления рисками в крупных производственных комплексах. Предлагается подход к созданию информационно-аналитической системы (ИАС) поддержки риск-ориентированного управления, основанный на использовании возможностей современных ИТ, инструментальных средств разработки профессионально-ориентированных систем и методов интеллектуального анализа данных.

Понятие риска и современное состояние проблемы управления рисками

Отсутствие однозначного понимания сущности риска на настоящий момент объясняется, в частности, многоаспектностью этого явления, отсутствием регулирования его законодательством. Это обусловило существование множества различных определений понятия «риск», представленных в литературе, в зависимости от целей исследования и точки зрения его автора [8].

Анализ предлагаемых в современных публикациях определений позволяет сделать заключение о тесной связи понятий «риск», «вероятность» и «неопределенность»: *в основе рыночного риска лежит вероятностная природа рыночной деятельности и неопределенность ситуации при ее осуществлении*. Еще одна сторона риска – его *принадлежность какой-либо деятельности*. Даже если компания не осуществляет никаких действий, она все равно несет риски – риск неполученной прибыли и пр. Это следует из сущности предпринимательской деятельности.

В соответствии с моделью COSO ERM *риск – это возможность наступления события, которое окажет отрицательное воздействие на достижение поставленных целей организации* [24]. С нашей точки зрения, данное определение представляется наиболее корректным в отношении предпринимательских,

финансовых рисков, так как отражает целевую направленность деятельности коммерческой организации. Расширенное содержание понятия охватывает основополагающие концепции управления рисками организаций, предоставляя основу для управления рисками в различных типах организаций, в разных отраслях и секторах экономики, т.е. является *общеприменимым*.

Системный риск – это риск потерь, связанных с неблагоприятными изменениями на рынке в целом, вызванных «эффектом домино» на финансовом рынке (в случае, если кризис одного или группы финансовых институтов/компаний реального сектора, кризис сегмента рынка или системы расчетов передается в расширяющемся объеме через пересекающиеся обязательства на другие группы финансовых институтов/компаний реального сектора, сегменты рынка и системы расчетов, постепенно охватывая всё расширяющуюся область рынка, или кризис финансового рынка одной страны или группы стран передается на другую страну) или *кризисом доверия* среди инвесторов, создающим ситуацию общей неликвидности на рынке [20].

Анализ многочисленных определений риска позволяет выделить следующие основные характеристики *рисковой ситуации*: наличие альтернативных решений; случайный характер события, который определяет, какой из возможных исходов реализуется на практике (наличие неопределенности); известны или определяемы вероятности исходов и ожидаемые результаты, вероятность возникновения убытков, вероятность получения дополнительной прибыли [7].

В соответствии с анализом 1 500 американских компаний, проведенным Э. Метриком, П. Гомперсом и Дж. Ишии, эффективная организация *систем управления рисками* (СУР) приводит к значительному росту стоимости этих компаний (примерно на 10% в год) [6]. Вместе с тем, существует ряд факторов, которые сдерживают применение СУР на отечественных предприятиях, о чем свидетельствуют результаты опроса [2], проведенного в 2007 г. русским обществом управления рисками среди российских компаний (табл. 1.).

Таблица 1. Факторы, сдерживающие применение систем управления рисками в России

Факторы	Доля респондентов, %
Отсутствие квалифицированных специалистов	65,52
Отсутствие структурированной информации для анализа и мониторинга риска	58,62
Отсутствие понимания у руководства организации потребности в системе	39,66
Создание системы не входит в стратегические цели организации	37,93
Отсутствие национальных стандартов управления рисками	34,48
Высокие затраты на внедрение системы	22,41

Как видно из таблицы, одним из весомых факторов является *недостаток информации для анализа, оценки и мониторинга рисков*. Проведенный анализ различных источников и результаты обследований ряда крупных промышленных предприятий Пермского края также подтверждают недостаточность научно-методического, информационного обеспечения внедрения и функционирования СУР в крупных производственных системах, региональном управлении этими системами. На это указывает перечень выявленных в ходе обследования наиболее существенных и часто встречающихся недостатков СУР коммерческих организаций (табл. 2).

Отечественные предприятия все чаще обращаются к зарубежному опыту разработки и внедрения СУР, а именно – к риск-ориентированной аудиторской концепции управления риском, известной как модель COSO (*Committee of Sponsoring Organizations of the Treadway commission*) [3]. Для организации системы управления финансовыми рисками была разработана методология «Управление рисками организации – интегрированная модель (COSO ERM)» [24], которая стала своеобразным развитием общего документа COSO, установившим основные понятия и логику организации системы управления финансовыми рисками на предприятии. В соответствии с методологией COSO ERM *система управления рисками* – это процесс, осуществляемый советом директоров, менеджерами и другими сотрудниками, который

начинается при разработке стратегии и затрагивает всю деятельность организации. Он направлен на выявление потенциальных событий, которые могут влиять на организацию, управление связанными с этими событиями рисками, а также контроль неперевышения «риск-аппетита» организации и предоставлением разумной уверенности в достижении целей организации. *Управление рисками* в соответствии с этой методологией представляет собой *непрерывный процесс, охватывающий всю организацию, осуществляемый сотрудниками на всех уровнях, используемый при разработке и формировании стратегии, применяемый на каждом уровне и в каждом подразделении организации*, включающий анализ портфеля рисков на уровне организации, нацеленный на выявление потенциальных событий, способных оказать влияние на организацию и управление рисками таким образом, чтобы они не превышали «риск-аппетит», т.е. степень риска, которую организация в целом считает для себя допустимой в процессе достижения поставленных целей.

Таблица 2. Типичные недостатки СУР в коммерческих организациях

Элементы СУР	Выявленный риск
Выявление рисков	Процедуры управления рисками не обеспечивают полноту выявления рисков
Оценка рисков	Процесс оценки рисков не позволяет своевременно и с необходимой точностью оценить влияние рисков на достижение целей организации. Система документирования рисков по направлениям деятельности организации не унифицирована
Реагирование на риски	Процесс реагирования на риски недостаточно увязан с бюджетным процессом и, как следствие, не позволяет снижать риски до приемлемого уровня
Мониторинг рисков	Отсутствует текущий контроль системы управления риском по основным показателям ее эффективности. Периодический мониторинг СУФР не проводится или проводится нерегулярно
СУФР дочерних предприятий	Отсутствует единая система управления финансовыми рисками головной организации/управляющей компании и дочерних предприятий

В рамках системного подхода к управлению рисками, который лежит в основе COSO ERM, предполагается *непрерывный анализ рисков* на предприятии, выделяются *компоненты процесса управления рисками* организации (компоненты СУР): внутренняя среда; постановка целей; выявление потенциальных событий; оценка рисков; реагирование на риск; контрольные процедуры; информация и коммуникации; мониторинг. Это – основа разрабатываемых моделей управления рисками.

В процессе управления рисками применяется четыре основных *способа реагирования на риск*:

- избежание риска – прекращение деятельности, ведущей к риску;
- принятие риска – не предпринимаются никаких действий для того, чтобы снизить вероятность влияния события (применяется к рискам, чье прогнозное влияние на финансовые результаты деятельности предприятия незначительно);
- снижение риска – действия по уменьшению влияния риска (требует принятия решений, касающихся организации деятельности);
- передача риска – уменьшение влияния риска за счет переноса или иного распределения части риска (страхование и др. способы).

Существуют различные технологии в области риск-менеджмента, предлагаемые крупными ИТ-компаниями, но основные области внедрения таких систем чаще связаны с финансовой деятельностью (банки, страховые компании). Это обусловлено тем, что финансовые риски поддаются математическому расчету и прогнозу с определенной долей вероятности. Однако в последние годы риск-менеджмент привлекает все большее внимание специалистов в разных сферах управления, так как риски увеличиваются параллельно с ростом сложности управляемых систем. Вместе с тем, разработанные за рубежом инструменты и технологии управления риском не всегда могут быть полностью и однозначно использованы отечественными предприятиями, поскольку они ориентированы на иные, существенно отличающиеся от российских, условия хозяйствования.

Наряду с рассмотренными выше проблемами, действует еще ряд факторов, обуславливающих *актуальность и необходимость разработки и внедрения комплексной информационно-аналитической СУР (ИАСУР)* в региональных производственных комплексах.

Во-первых, это *асимметрия информации* о рисках, возникающая в силу их множественности, неопределенности наступления рисков, проистекающей из самой его сущности и закрытости большей части информации о рисках и системах управления ими в хозяйствующих субъектах, финансовых институтах. Усиление предпринимательского интереса к сохранению коммерческой информации о структуре и способах управления рисками в условиях нестабильности экономики создаёт существенные угрозы устойчивого развития отдельных территорий, государств и всего мирового сообщества.

Другим фактором, определяющим объективную необходимость создания комплексной ИАСУР, является *связанность рисков в социальных системах*, включая региональные производственные комплексы, как правило, интегрированные в международные рынки, а также *способность рисков к концентрации и накоплению* вследствие неэффективного управления и наличия взаимосвязи и взаимной зависимости различных рисков. При отсутствии эффективных механизмов и моделей более раннего выявления, идентификации признаков системных рисков (теории «черного лебедя», «королей драконов» и др.) это несет существенные угрозы устойчивому функционированию РПК.

Проблема обостряется тем, что вслед за усложнением социально-экономических, производственных процессов и систем управления, *усложняется и процесс управления рисками* в них. Таким образом, *происходит постоянное пополнение перечня уже известных рисков* при развитии процессов специализации и разделения общественного труда, диверсификации направлений деятельности хозяйствующих субъектов. Решение указанной проблемы невозможно без объединения усилий ученых и практиков различных отраслей и сфер деятельности при формировании перечней рисков, оценке их последствий и разработке способов и инструментов управления ими.

Сохранение асимметрии информации по управлению рисками в силу распыления сил и отсутствия консолидации знаний различных участников рынка отражается на качестве решений, принимаемых менеджментом хозяйствующих субъектов и государственных органов, ведет к неэффективной работе рынка и неэффективному размещению ресурсов в результате проблем неблагоприятного отбора (*adverse selection*), рисков недобросовестности (*moral hazard*), дорогостоящей верификации состояния (*costly state verification*). В конечном счете, все это также способствует накоплению рисков, создает угрозы устойчивому развитию субъектов хозяйствования, а также развитию территорий.

Как следствие сказанного выше, на настоящий момент во многих российских коммерческих организациях применяется лишь *фрагментарная система управления рисками*, которая, во-первых, не позволяет оперативно отслеживать все наиболее существенные риски, а во-вторых, не позволяет эффективно передавать информацию о рисках всем заинтересованным сторонам. Острой проблемой остается *реактивный характер управления рисками*. Информация по рискам хранится в разрозненных системах. В результате компании не имеют возможности своевременной идентификации критических рисков, их анализа и грамотного реагирования на риски с использованием результатов комплексного анализа.

Для решения данной проблемы предлагается использовать *системный подход в управлении рисками*, внедрять процедуры *проактивного мониторинга* в существующие бизнес-процессы и стратегии. Средства управления рисками должны быть интегрированы с информационными системами (ИС), методика системного управления рисками должна перерасти в единую ИТ-стратегию компании [19]. Это делает особенно актуальной разработку концепции, методологии управления рисками на базе современного информационно-технологического обеспечения.

Локальные системы мониторинга рисков и тенденции интеграции

В настоящее время в России существует множество центров и локальных систем мониторинга рисков, у

каждого из которых имеются свои цели и методики. Среди них наиболее известны следующие:

1. Системы *мониторинга устойчивости банковского сектора*. Эти системы достаточно детально проработаны, их характеристики представлены в литературе и нормативных документах. При их реализации используются инструменты мониторинга, которые могут быть полезны при разработке систем мониторинга и в других секторах экономики (традиционные методы статистического наблюдения, прогнозирования, анализа рядов динамики, методы математической статистики, методы экспертных оценок, относительных оценочных показателей, агрегатных индексов и др.).

2. Текущий и периодический *мониторинг рисков нарушения законодательства, мошенничества, реализации коррупционных схем и неэффективного использования бюджетных средств*, осуществляемый органами государственного, регионального и муниципального контроля (в том числе и. финансового контроля, экологического и природоохранного контроля, пожарной безопасности, контроля чрезвычайных ситуаций и т.д.).

3. Системы *мониторинга финансово-экономического состояния предприятий*, реализующие методы диагностики, оценки и прогноза финансового состояния предприятий с целью предотвращения кризисных ситуаций и повышения эффективности экономической деятельности предприятий и финансовых институтов, обеспечения их безопасности и устойчивого развития [5]:

- мониторинг 150 тыс. крупных и средних предприятий Федеральной службой государственной статистики с публикацией ежегодных отчетов, содержащих в том числе аналитические показатели доли имобилизованных активов в имуществе, коэффициенты обеспеченности запасов и затрат собственными средствами, текущей и абсолютной ликвидности, сравнение уровней рентабельности производства (<http://www.gks.ru/wps/portal>);
- мониторинг финансово-экономического состояния промышленных предприятий, осуществляемый с 2004 г. территориальными подразделениями Минэкономразвития РФ, региональные ИАС мониторинга финансовой устойчивости территорий (<http://ias.perm.ru/ias-gpk>);
- внешний мониторинг Центрального Банка РФ (с 2000 г.) финансового состояния предприятий с целью осуществления анализа важнейших тенденций на разных уровнях территориального управления, снижения риска банков при работе с предприятиями (Положение о проведении мониторинга предприятий Банком России №186-П от 19 марта 2002 года);
- мониторинг правоохранительных органов, органов государственной безопасности и др.

4. *Комплексная система мониторинга рисков финансового рынка*, внедряемая в настоящее время ФСФР (http://www.lin.ru/news_item.htm?id=5532123061415052721).

5. Текущий и периодический *мониторинг рисков корпоративного управления и отдельных бизнес-процессов* (прежде всего процесса составления финансовой отчетности), эффективности СУР, осуществляемый службами риск-менеджмента, внутреннего контроля и аудита корпоративных хозяйствующих субъектов либо привлеченными аудиторскими и консалтинговыми компаниями. В качестве основной методологии используется Методология COSO, которая получает все большее распространение среди российских ученых и практиков благодаря работе на российском рынке крупных зарубежных аудиторских компаний [23].

6. *Обследования и оценки, проводимые различными научно-исследовательскими институтами и центрами по наиболее важным конъюнктурным показателям*, которые, впрочем, являются достаточно общими и не предоставляют полной информации об устойчивости финансового положения промышленных предприятий и наиболее существенных рисках (<http://www.imemo.ru/barom/>; <http://www.iet.ru/>; http://www.hse.ru/temp/2005/09_05_research.shtml и др.).

В рамках функционирования локальных систем контроля над рисками происходят очень важные процессы накопления эмпирической информации о рисках, их последствиях, методах управления ими, перечней (реестров, библиотек) рисков, форм их документирования, инструментария оценки и т.д.

Следствием независимого функционирования множества локальных и разрозненных информационно-

аналитических подсистем контроля рисков является:

1. Отсутствие преемственности при описании бизнес-процессов хозяйствующих субъектов для разных целей и разных категорий пользователей с такими негативными последствиями как

- изоляция и несогласованное функционирование систем риск-менеджмента и процессов разработки, внедрения процедур контроля, текущего и периодического мониторинга;
- дублирование и отсутствие унификации документирования процессов в различных подсистемах управления;
- проведение текущего и периодического мониторинга эффективности процедур контроля внутренними аудиторскими и владельцами бизнес-процессов без учета зон повышенного риска.

2. Нерациональные и значительные по объему непроизводительные затраты, направленные на поиск, сбор и обобщение разрозненной информации для разработки, внедрения и поддержания в актуальном состоянии СУР.

В результате увеличиваются затраты на разработку и поддержание СУР, резко снижается эффективность осуществления контроля над рисками.

Анализ локальных систем мониторинга рисков показывает, что ни один из центров мониторинга не решает задачу реализации государственной политики в области контроля над рисками с целью обеспечения устойчивого развития регионов, важнейших РПК. В настоящий момент отсутствует единая методика и концепция проведения мониторинга рисков регионов, РПК, который позволил бы эффективно и своевременно выявлять потенциально опасные для развития региона риски, оценить их и предложить набор возможных мер по оптимизации рисков для принятия управленческих решений, проведения мероприятий по профилактике кризисных ситуаций.

XXI век положил начало новым интегративным процессам и тенденциям в области управления рисками:

- работа по снижению уровня международной организованной преступности, коррупции и противодействию легализации преступных доходов и терроризму, основанная на тесном международном сотрудничестве [4];
- тенденции по объединению усилий независимых органов финансового контроля государственного и корпоративного секторов экономики в области методологии разработки и оценки эффективности систем контроля, подготовки кадров по финансовому контролю, начало которой было положено соглашением между Контрольно-счетной палатой РФ и международным институтом внутренних аудиторов в 2010 г. (<http://www.iaa-ru.ru>; <http://www.iaa-ru.ru/>);
- работы по совершенствованию методологии корпоративного управления, управления финансовыми рисками, проводимые крупнейшими аудиторскими и консалтинговыми компаниями, рейтинговыми агентствами в тесном сотрудничестве с крупными корпорациями, саморегулируемыми организациями [24].

Развитие процессов интеграции и глобализации способствует аккумуляции опыта различных институтов управления рисками: банков, страховых компаний, государственных и муниципальных органов, корпоративных образований и отдельных предприятий в рамках развития межотраслевой, межрегиональной интеграции. Начавшиеся процессы интеграции, безусловно, требуют соответствующей информационной поддержки.

Архитектура и методы реализации информационно-аналитической системы поддержки процессов управления рисками

Проведенный анализ показывает следующие проблемы, препятствующие созданию эффективных систем управления рисками РПК:

- *Отсутствие единой устоявшейся терминологии* в данной области: даже в рамках одной отрасли зачастую, с одной стороны, используются различные определения одних и тех же

понятий, а с другой, применяются различные термины для описания одних и тех же объектов и их характеристик.

- *Недостаточность научно-методического, информационного обеспечения* реализации систем управления рисками в крупных производственных системах, региональном управлении.
- *Асимметрия информации* о рисках в РПК, возникающая в силу их множественности и закрытости большей части информации о рисках и системах управления ими в хозяйствующих субъектах, финансовых институтах: недостаток информации при ее дублировании и избыточности в локальных системах.
- *Трудоемкость выявления ключевых показателей*, причинно-следственных связей между финансовыми и нефинансовыми, ключевыми и оперативными показателями в привязке к стратегическим целям компании и бизнес-процессам, зонам риска РПК.

Представленная в данной работе ИАС не является СУР, реализующей все перечисленные выше компоненты. В данном случае поставлена задача создания средств поддержки работы аналитиков, экспертов, профессиональными задачами которых является разработка компонентов СУР и их оценка, выработка рекомендаций по усовершенствованию СУР, их оптимизации.

Информационно-аналитическая система, ориентированная на решение перечисленных задач, должна включать следующие компоненты (рис. 1):

- Компонент *интеллектуального поиска и анализа документов*, основанные на использовании онтологических моделей предметных областей [10, 11, 12], создаваемых пользователями системы с использованием редактора моделей, позволяющего строить модели в привычных для пользователей системы терминах предметной области (для работы с моделями используются возможности DSL-инструментария).
- Компонент *интеллектуального анализа данных*, предназначенный для анализа данных о состоянии бизнес-систем, предприятий РПК с целью выявления значимых с точки зрения их деятельности фактов, идентификации признаков системных рисков и пр. Реализация данного компонента основывается на методах анализа многомерных данных, а также теории исчисления фактов [15, 16, 17].
- Компонент *анализа моделей управления рисками* – центральный компонент ИАС. Данный компонент позволяет пользователям создавать и анализировать модели, которые могут служить основой для разработки СУР, – типовые модели, в которых обобщается опыт создания и использования СУР в различных областях. При создании моделей эксперт имеет возможность не только использовать найденные источники информации, относящиеся к соответствующей предметной области, загруженные в библиотеку документов, но и опираться на возможности средств интеллектуального анализа данных, получаемых их различных гетерогенных источников.
- *Языковой инструментарий* с редактором моделей – компонент, позволяющий создавать предметно-ориентированные языки (DSL – Domain Specific Languages) для разработки моделей [13, 14]. DSL-инструментарий – основа для разработки динамически адаптируемой профессионально ориентированной системы, функционирование которой описывается созданными моделями.

Схема взаимодействия компонентов информационно-аналитической системы показана на рис. 1.

В системе необходимо реализовать инструментарий, обеспечивающий автоматизацию трудоемких операций по поиску и анализу данных, разработке моделей управления рисками и их апробации и пр. Многомерная классификация и удобная каталогизация ресурсов, наличие средств навигации, настраиваемых в соответствии с запросами пользователей, – еще одно требование к системе.

Основа системы – *предметно-ориентированное моделирование* (DSM – Domain Specific Modelling) [25]. Работа строится на использовании моделей различных типов, создаваемых экспертами с помощью реактора моделей. Для разработки моделей в состав системы включен языковой инструментарий: с его

помощью экспертами создаются *предметно-ориентированные языки* (DSL), позволяющие пользователям работать с моделями в привычных для них терминах предметной области. Особенность *DSL-инструментария* – возможность динамической настройки, разработки новых языков на основе ранее созданных, что дает возможность экспертам, построившим базовый язык предметной области (метаязык), постепенно уточнять и расширять его возможности, создавая новые языки-модели, отражающие специфику новых областей применения системы. Разрабатываемые онтологические модели дают возможность определить все используемые понятия предметной области и связи между ними, описать документы, относящиеся к определенной области, и т.д. Реализация языкового инструментария основана на использовании мета- и гиперграфов, графовых грамматик и графовых трансформаций [13, 14].

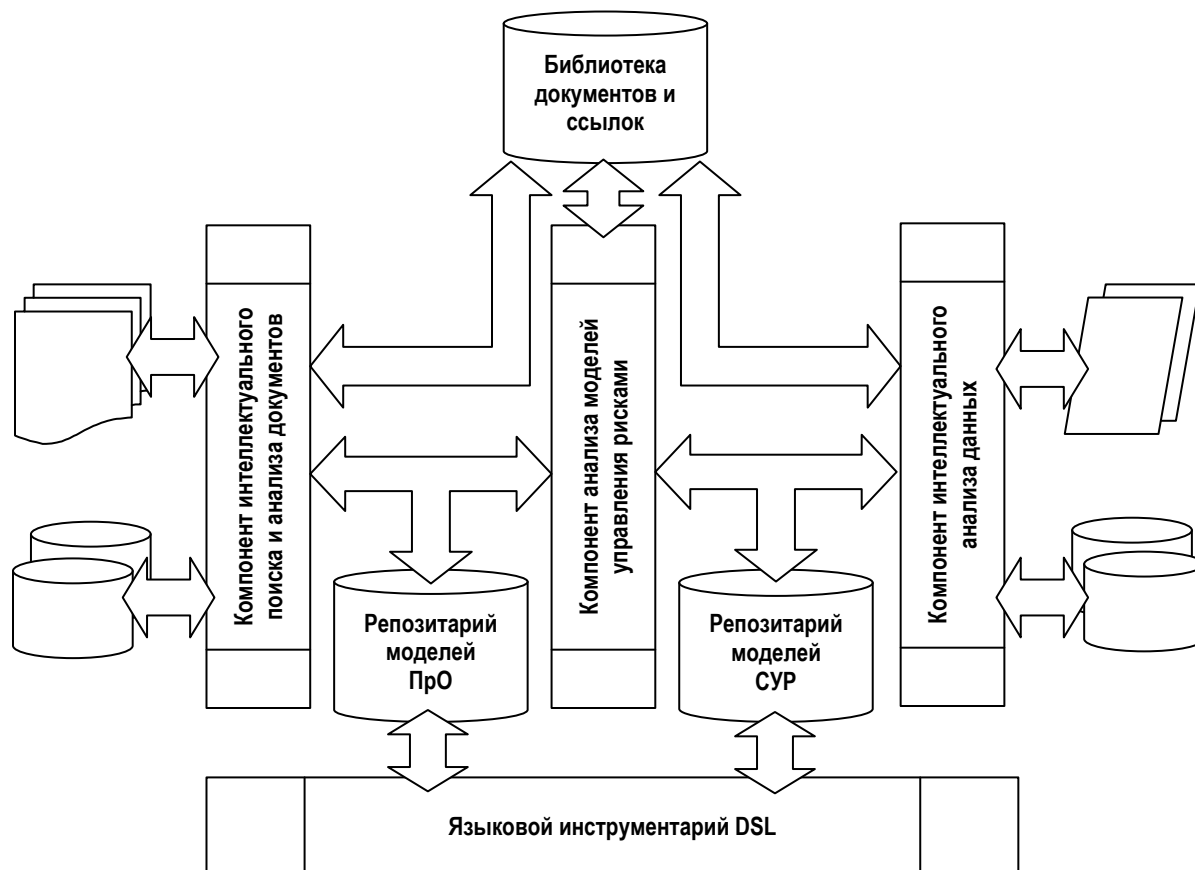


Рис. 1. Архитектура информационно-аналитической системы поддержки управления рисками

Разработанные пользователями модели – основа функционирования *подсистемы интеллектуального поиска и анализа документов*. Результат, получаемый пользователями, – система документов, отвечающих онтологическим моделям, разработанным пользователями, описывающих состояние соответствующей предметной области. Найденные документы каталогизируются в библиотеке с сохранением ссылок, отражающих существующие между ними связи, обеспечивающих удобную навигацию по найденным источникам. Выполняется автоматическое реферирование документов. Средства поиска и анализа документов позволяют снизить трудоемкость поиска информации о предметной области (нормативно-правовых актов, регламентирующих деятельности РПК, исследовательских работ и пр.). Найденная информация может служить основой для уточнения ранее построенных моделей, их расширения, настройки на меняющиеся условия.

Основная цель создания системы – обобщение опыта разработки и эксплуатации СУР, поддержка исследователей, занимающихся проблемами управления рисками. Центральный компонент информационно-аналитической системы – *подсистема анализа моделей управления рисками*. Как было отмечено выше, риски всегда связаны с бизнес-процессами, поэтому основа создания моделей – средства формализации описания деятельности компаний, бизнес-процессов, основанные на использовании соответствующих языков – языков описания бизнес-процессов [21]. На основе существующих языков с помощью языкового инструментария эксперты могут создавать новые языки, дополненные средствами описания рисков, связанных с бизнес-процессами. Система позволяет разработать языки для различных областей деятельности, отражающие их специфику, на основе базового метаязыка. Созданные с использованием этих языков модели сохраняются в репозитории моделей управления рисками и могут использоваться в качестве типовых моделей пользователями системы. Средства анализа моделей основываются на традиционных математических методах (методах теории графов, статистических методах и пр.).

Заключение

В работе представлен проект информационно-аналитической системы поддержки исследований проблемы управления рисками, описаны подходы к реализации ее компонентов, основанные на ранее полученных результатах. Реализация ИАС позволит создать ядро «саморазвивающейся» системы, на базе которой могут быть интегрированы усилия экспертов, работающих в данной области. Созданный ресурс позволит объединить усилия ученых и практиков по созданию эффективных СУР.

Благодарности

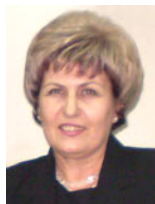
The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Библиографический список

1. Бадалова А.Г. Система управления рисками: методология, организационно-информационное обеспечение, эффективность внедрения. Монография. – М.: «Станкин», «ЯНУС-К», 2007.
2. Вашакмаждзе Т., Гребенюк А. Как финансовому директору управлять рисками // Финансовый директор.– 2007.– № 7-8. – С. 20.
3. Внутренний контроль: интегрированный подход: Доклад Комитета спонсорских организаций Комиссии Тридуэя (the Committee of sponsoring organizations of the Treadway Commission's internal control – integrated framework) [Электронный ресурс]. – 1992. – Режим доступа: <http://www.coso.org>.
4. Гамза В.А., Ткачук И.Б. Безопасность банковской деятельности: Учебник. – М.: Маркет ДС, 2006.
5. Давыдова Л.В., Гордина В.В. Этапы осуществления мониторинга промышленных предприятий в Российской Федерации [Электронный ресурс] // Управление общественными и экономическими системами. – 2010. – Режим доступа: <http://bali.ostu.ru/umc/z022005.php>.
6. Дубовицкая Е. Риски корпоративного управления // Проблемы теории и практики управления. – 2007. – № 12. С. 84.
7. Ильин В.В, Сердюков Н.А. Системный подход к оценке финансовых рисков // Финансы. – 2008. – № 1. – С. 68.
8. Кутергина Г.В., Модорский А.В. Анализ эффективности систем управления финансовыми рисками коммерческой организации // Аудит и финансовый анализ. – 2010. – № 3. С. 149-162.
9. Ланин В.В. Методы и средства решения задач информационного поиска для системы поддержки научных исследований // Инновационное развитие регионов: методы оценки и поддержка исследований: межвуз. сб. науч. статей / Перм. гос. ун-т. – Пермь, 2009. С. 80-88.
10. Ланин В.В. Решение задач информационного поиска для исследовательского портала на основе агентного и онтологического подходов // Инновационное развитие регионов: методы оценки и поддержка исследований: межвуз. сб. науч. статей / Перм. гос. ун-т. – Пермь, 2009. С. 89-96.
11. Ланин В. Онтологии как основа функционирования систем обработки электронных документов // Материалы Всероссийской конференции с международным участием «Знания-Онтологии-Теории». – Новосибирск, 2009. Т. 2. С. 173-177.

12. Ланин В.В. Построение системы взаимосвязанных документов на основе средств интеллектуального анализа // Труды Конгресса по интеллектуальным системам и информационным технологиям (AIS-IT'09) / М.: Физматлит, 2009, Т.1. С.426-431.
13. Лядова Л.Н., Сухов А.О. Визуальные языки и языковые инструментари: методы и средства реализации // Труды Конгресса по интеллектуальным системам и информационным технологиям «AIS-IT'10» Научное издание в 4-х томах. Т. 1. – М.: Физматлит, 2010. С. 374-382.
14. Лядова Л.Н., Сухов А.О. Языковой инструментарий системы MetaLanguage // Математика программных систем: межвуз. сб. науч. ст. / Перм. гос. ун-т. – Пермь, 2008. С. 41-52.
15. Мальцев П. Моделирование многомерных данных в системе METAS BI-PLATFORM // Advanced Studies in Software and Knowledge Engineering: International Book Series / Sofia, 2008. С. 173-180.
16. Мальцев П.А. Моделирование и анализ фактов и связей между ними // Natural and Artificial Intelligence: International Book Series / Sofia, 2010. С. 194-199.
17. Мальцев П.А., Лядова Л.Н. Формализация многомерной модели данных // Математика программных систем: Межвузовский сб. науч. тр. / Перм. ун-т. Пермь, 2006. С. 74-87.
18. О стратегии национальной безопасности РФ до 2020 г.: указ Президента Российской Федерации от 12 мая 2009 г. № 537 // СПС Консультант плюс [Электронный ресурс] / ЗАО «ТелекомПлюс». – Пермь, 2011. 06.03.2011. П. 63.
19. Полтев А.В. Системный подход в управлении рисками организаций, внедряющих крупную учетную систему класса ERP // Труды 53-й научной конференции МФТИ «Современные проблемы фундаментальных и прикладных наук». Часть XI. Информационные бизнес-системы. – М.: МФТИ, 2010. С. 24-25.
20. Риски финансового кризиса в России: факторы, сценарии, политика противодействия: Национальный доклад финансовой академии при Правительстве РФ / Я.М. Миркин; Институт финансовых рынков и прикладной экономики Финансовой академии при Правительстве РФ. – Москва, 2008.
21. Рычков А.Ю. Автоматизация управления бизнес-процессами на основе языка GBPR // Intelligent Support of Decision Making / International Book Series “Information Science & Computing”, Number 10. Sofia, Bulgaria, 2009.
22. Савич А.В., Дианова Н.Ю. Управление бизнес-процессами – основа эффективного управления компанией // Труды 53-й научной конференции МФТИ «Современные проблемы фундаментальных и прикладных наук». Часть XI. Информационные бизнес-системы. – М.: МФТИ, 2010. С. 103-105.
23. Сонин А.М. Внутренний аудит: современный подход [Текст] / А.М. Сонин. – М.: Финансы и статистика, 2007.
24. Управление рисками организаций. Интегрированная модель: Свод общих положений. – Комитет спонсорских организаций Комиссии Тридуэя (COSO) [Электронный ресурс] / Институт внутренних аудиторов, 2004. – Режим доступа: http://www.iaa-ru.ru/files/documents/COSOFramework_1.pdf.
25. Lyadova L. Metamodeling and Multilevel Metadata as a Basis of Technology that is Intended for Development of Adaptable Information Systems // International Journal “Information Technologies & Knowledge”. Volume 3 / 2009, Number 1. – Sofia. P. 67-81.

Сведения об авторах



Галина Кутергина – Пермский государственный университет, доцент кафедры финансов, кредита и биржевого дела, Россия, г. Пермь, 614990, ул. Букирева, д. 15; e-mail: GKutergina@psu.ru.
Major Fields of Scientific Research: системы управления финансовыми рисками; внутрифирменный финансовый контроль.



Людмила Лядова – Пермский филиал Национального исследовательского университета «Высшая школа экономики», доцент кафедры информационных технологий в бизнесе; Россия, г. Пермь, 614070, ул. Студенческая, д. 38; e-mail: LNLyadova@mail.ru.
Major Fields of Scientific Research: метамоделирование; технология DSM; CASE-средства; языковые инструментари; предметно-ориентированные языки, DSL.



Наталья Фролова – Пермский государственный университет, доцент кафедры информационных систем и математических методов в экономике, Россия, г. Пермь, 614990, ул. Букирева, д. 15; e-mail: nvf_psu@mail.ru.
Major Fields of Scientific Research: графовые модели бизнес-процессов и систем; графовые грамматики; аналитические системы.

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ДИСЦИПЛИНЫ «ЭЛЕКТРИЧЕСКИЕ МАШИНЫ» / ИНТЕРНЕТ-РЕАЛИЗАЦИЯ /

Евгений Забудский

Аннотация: Разработан учебно-методический комплекс дисциплины «Электрические машины». Комплекс включает программу дисциплины, учебные пособия с грифом, описание и фотографии разработанных лабораторных стендов, исходные тексты компьютерных программ, техническое задание и методические материалы по курсовой работе, мультимедийные лекции, слайд-фильмы, гиперссылки на Интернет-ресурсы, etc. Комплекс представлен в интернете (<http://zei.narod.ru/ind.html>).

Ключевые слова: Электрическая машина, трансформатор, магнитное поле, Интернет, веб-сайт, комплекс.

ACM Classification Keywords: K.3.1 Computer Uses in Education: Distance learning.

Введение

Электрические машины и трансформаторы представляют основную и важнейшую часть силового электрооборудования. Поэтому глубокое изучение их теории и практики является неотъемлемой составляющей процесса подготовки инженеров электриков. Спецификой устройств, затрудняющей изучение, является сложность представления и понимания «рабочего тела» устройств – магнитного поля. Кроме того при относительной простоте конструкций, их иллюстрация на лекциях занимает много времени, и с помощью только традиционных средств не всегда является эффективной. Для устранения указанных проблем учебно-методический комплекс дисциплины среди прочего включает разработанные мультимедийные приложения: компьютерные звуковые фильмы, в которых представлены в деталях конструкции устройств, дана интерпретация магнитного поля в пространственно-временном континууме, изложены элементы теории. Использование комплекса в учебном процессе преподавателями и студентами повышает эффективность изложения и усвоения дисциплины «Электрические машины».

Учебно-методический комплекс

На рис. 1 представлена web-страница с расположенным на ней учебно-методическим комплексом (УМК) дисциплины «Электрические машины» (URL-адрес – <http://zei.narod.ru/index1.html>).

Далее приводится краткое описание и иллюстрации (рис. 2 ... рис. 5) некоторых методических мультимедийных материалов, к которым имеется доступ посредством гиперссылок на этой web-странице [1 ... 7].

Спроектированы и изготовлены стенды к лабораторным работам по разделам курса «Трансформаторы», «Синхронные машины» и «Машины постоянного тока» (см. п. 6 УМК). Реализован фронтальный метод проведения работ. Приведены фотографии и описание стендов (рис. 2). При домашней подготовке к занятиям, имея доступ к Интернету, студенты могут умозрительно и виртуально собрать соответствующую электрическую схему (см., например, [4] и <http://zei.narod.ru/CM.html>).

Опубликованы в электронном виде (и в твердом варианте) учебные пособия по указанным выше разделам курса (см. п.п. 9.1...9.3 УМК, <http://zei.narod.ru/index1.html>). Приведены описания лабораторных работ, схемы, нормативные материалы, основы теории, etc. (рис. 3). С сайта возможно скачать (public domain) эти пособия в полном объеме.

Учебно-методический комплекс дисциплины Электрические машины (фото)

1	Забудский Е.И. Программа дисциплины Электрические машины . Рекомендована Министерством образования и науки России для направления подготовки дипломированного специалиста 660300 - "Агроинженерия" (специальность 311400 - Электрификация и автоматизация сельского хозяйства)
2	Забудский Е.И. Электрические машины. Ч. 1. Трансформаторы . Учебное пособие для вузов. - Москва: МГАУ, 2002. – 166 с. 2.1. Форма отчета по Лабораторной работе (образец оформления; на примере Работы №1) 2.2. Паскаль-программа "Расчет характеристик холостого хода трансформатора" (обработка данных эксперимента)
3	Забудский Е.И. Электрические машины. Ч. 2. Асинхронные машины . Учебное пособие для вузов. - Москва: МГАУ, 2002. – 100 с. (оригинал-макет) 3.1. Паскаль-программа "Расчет характеристик холостого хода асинхронного двигателя" (графическая интерпретация; обработка данных эксперимента)
4	Забудский Е.И. Электрические машины. Ч. 3. Синхронные машины . Учебное пособие для вузов. - Москва: МГАУ, 2008. – 195 с. 4.1. Паскаль-программа "Расчет экспериментальных значений величин в системе относительных единиц" (графическая интерпретация)
5	Забудский Е.И. Электрические машины. Ч. 4. Машины постоянного тока . Учебное пособие для вузов. - Москва: МГАУ, 2004. – 100 с. (оригинал-макет) 5.1. Паскаль-программа "Расчет рабочих характеристик двигателя постоянного тока с параллельным возбуждением" (графическая интерпретация; обработка данных эксперимента)
6	Информация о лабораториях Электрических машин: тексты лабораторных работ и фото стендов 6.1. Стенд для проведения лабораторных работ по разделу "Трансформаторы" (фото и описание) 6.2. Стенды для проведения лабораторных работ по разделу "Машины постоянного тока" (фото и описание) 6.3. Стенд для проведения лабораторных работ по разделу "Синхронные машины" (фото и описание) 6.4. Стенды для проведения лабораторных работ по разделу "Асинхронные машины" (стенды в стадии разработки)
7	Техническое задание и Методические рекомендации к курсовой работе "Расчет асинхронного двигателя с короткозамкнутым ротором" 7.1. Паскаль-программа "Расчет рабочих характеристик асинхронного двигателя" (к курсовой работе; руководство пользователю): загрузить файл RX_AD.rar . Файл сохранить в корне диска C (в rar-файле находятся файл RX_AD.exe и файл с исходными тест-данными RX_AD_dn.dat) 7.2. Каталоги Электрических машин и Трансформаторов , изготавливаемых заводами России, etc. (интернет-ресурс) 7.3. ФГУП Институт промышленного развития (Информэлектро) – Информационный центр России (интернет-ресурс) 7.4. Всероссийский институт научной и технической информации РАН (ВИНИТИ РАН) (интернет-ресурс) 7.5. Всероссийский научно-технический информационный центр (ВНТИЦентр) (интернет-ресурс) 7.6. Защита интеллектуальной собственности (РУПАТЕНТ) (интернет-ресурс) 7.7. Российский научно-технический центр по стандартизации, ... (СТАНДАРТИНФОРМ) (интернет-ресурс)
8	Экзаменационные вопросы по дисциплине Электрические машины: 1-й семестр (Тр и АМ) ; ... 2-й семестр (СМ, МПТ и ЭММ)
9	СТАТЬИ И ДРУГИЕ МАТЕРИАЛЫ , опубликованные на web-сайте ...

	9.1. Забудский Е.И. Электрические машины. Ч. 1. Трансформаторы : Учебное пособие для вузов.– М.: МГАУ, 2002. – 168 с. (полный текст пособия: скачайте ...)
	9.2. Забудский Е.И. Электрические машины. Ч. 3. Синхронные машины : Учебное пособие для вузов.– М.: МГАУ, 2008. – 196 с. (-, -, -: скачайте ...)
	9.3. Забудский Е.И. Электрические машины. Ч. 4. Машины постоянного тока : Учебное пособие для вузов.– М.: МГАУ, 2010. – 217 с. (-, -, -: скачайте ...)
	9.4. Забудский Е.И. Математическое моделирование управляемых электроэнергетических устройств : Учебное пособие для вузов. - Ульяновск: УлГТУ, 1998. – 120 с.
	9.5. Забудский Е.И. Анализ управляемых электроэнергетических устройств методом конечных элементов : Учебное пособие для вузов. - Москва: МГАУ, 1999. – 141 с.
	9.6. Забудский Е.И. Совмещенные регулируемые электромагнитные реакторы : Монография. - Москва: Энергоатомиздат, МГАУ, 2003. – 436 с.
10	Забудский Е.И. Комплекс математических моделей и компьютерных программ для расчета электромагнитных полей и режимов ферромагнитных устройств
11	Забудский Е.И. Пакеты компьютерных программ , зарегистрированных в Отраслевом фонде алгоритмов и программ НИИВО Министерства образования и науки России
12	Забудский Е.И. Компьютерный фильм “Геометрическая интерпретация результатов расчета магнитного поля в устройствах электромеханики” (сертификат №3/93)
Карта сайта профессора Е.И.Забудского ...	

Рис. 1. Web-страница с учебно-методическим комплексом (<http://zei.narod.ru/index1.html>)



Рис. 2. Приборная панель лабораторного стенда «Генератор постоянного тока» (<http://zei.narod.ru/MPT.html>)

Студенты имеют возможность просмотреть компьютерный слайд-фильм, в котором представлены результаты расчета магнитного поля, выполненного методом конечных элементов в сильнонасыщенном устройстве трансформаторного типа [8 ... 10] (фильм разработан в среде Графор, <http://zei.narod.ru/filmr/film.html>). В фильме наглядно представлены: изменение магнитных индукции, напряженности, проницаемости и векторного потенциала в пространственно-временном континууме; картина распределения силовых линий магнитного поля; результаты гармонического анализа, оформленные в виде гистограммы, etc. (рис. 4). В процессе работы студенты не только изучают характер распределения магнитного поля в пространстве и во времени, его зависимость от насыщения среды, от конфигурации области расчета и пр., но и развивают абстрактное мышление, позволяющее мысленно представить и воссоздать картину распределения поля в различных типах электрических машин и трансформаторов. Фильм имеет сертификат (<http://zei.narod.ru/15.html>).

На web-странице имеются также гиперссылки на файлы с исходными текстами паскаль-программ (см., например, п.7.1 УМК) на основе которых осуществляются: расчет рабочих характеристик асинхронного двигателя (http://zei.narod.ru/RX_AD_Pr.pdf), выполняемый в рамках курсового проекта. (рис. 5); обработка результатов эксперимента (http://zei.narod.ru/SM1_OE_Programma.pdf).

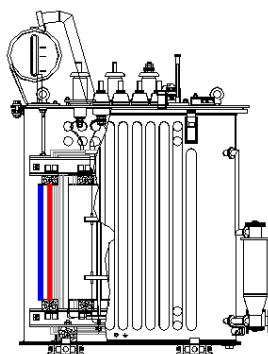
Кроме того на web-странице приведены гиперссылки на паскаль-программы в которых реализованы методы решения систем уравнений, моделирующих установившиеся симметричные, несимметричные и переходные процессы, происходящие в электрических машинах и трансформаторах (<http://zei.narod.ru/7.html>). К этим системам и методам их решения относятся: системы линейных алгебраических уравнений (метод Гаусса, <http://zei.narod.ru/Nuton.pdf>); системы нелинейных алгебраических и трансцендентных уравнений (метод Ньютона, <http://zei.narod.ru/Nuton.pdf>); системы обыкновенных дифференциальных уравнений (метод Рунге-Кутты с модификацией Фельберга, http://zei.narod.ru/SODU_2.pdf); системы дифференциальных уравнений в частных производных (метод конечных элементов, <http://zei.narod.ru/Up2.html>).

Ряд разработанных электронных мультимедиа-материалов в связи с их большим объемом не расположены в Интернете на web-странице УМК, но непосредственно демонстрируются с помощью ноутбука и мультимедийного проектора на лекциях и практических занятиях. К ним относятся:

- пять звуковых видеofilмов по конструкции и технологии сборки электрических машин в соответствии пятью разделами дисциплины: 1. Трансформаторы (рис. 6), 2. Асинхронные машины, 3. Синхронные машины, 4. Машины постоянного тока, 5. Тепловые и гидроэлектростанции (разработано в среде PowerPoint);
- мультимедиа-лекция по теме «Реакция якоря синхронной машины». В лекции представлен в динамике процесс влияния поля обмотки якоря на поле обмотки возбуждения при различном характере нагрузки, подключенной к обмотке якоря. Рис. 7 иллюстрирует это влияние при активной нагрузке (разработано в среде MicromediaFlash);
- видеоматериалы, иллюстрирующие в динамике условия создания вращающегося поля трехфазной обмоткой. Эти условия заключаются в следующем: 1) фазы обмотки должны быть сдвинуты в пространстве на 120 электрических градусов; 2) токи в этих фазах должны быть смещены во времени на 120 градусов. На рис. 8 представлено положение силовых линий вращающегося поля, созданного обмоткой статора, для двух моментов времени (разработано там же);
- видеоматериалы, иллюстрирующие в динамике процесс укладки и устройство якорной обмотки машины постоянного тока, а также выпрямление напряжения коллектором генератора, рис. 9 (там же).

Электрические машины

Часть 1. ТРАНСФОРМАТОРЫ

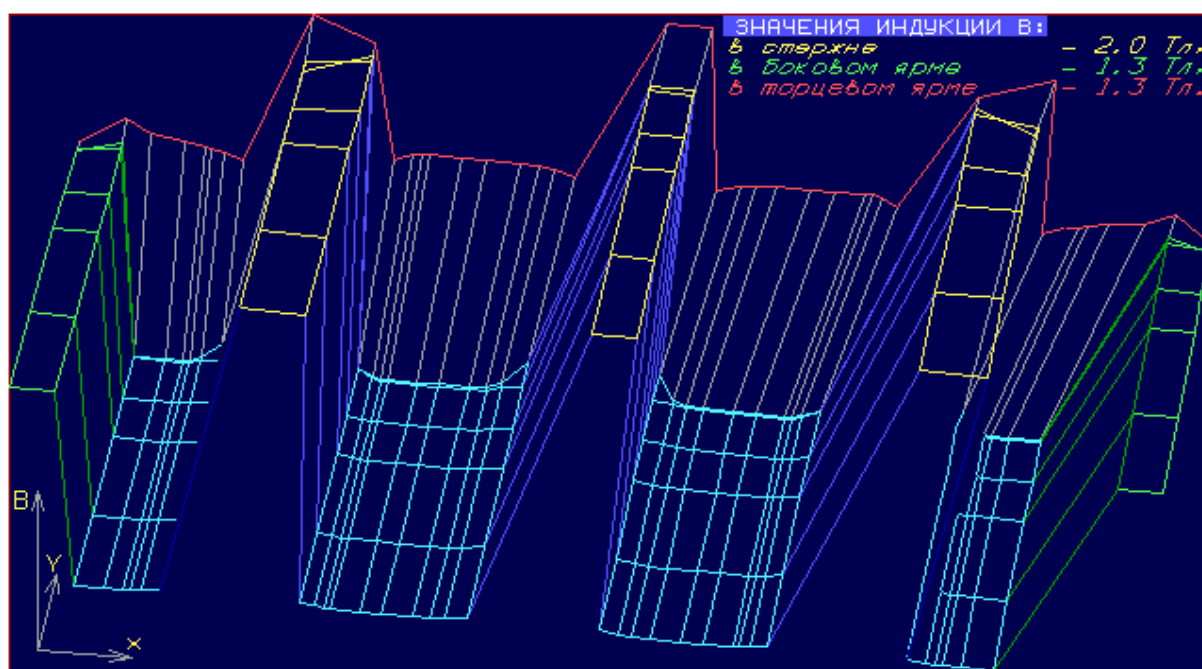


Содержание

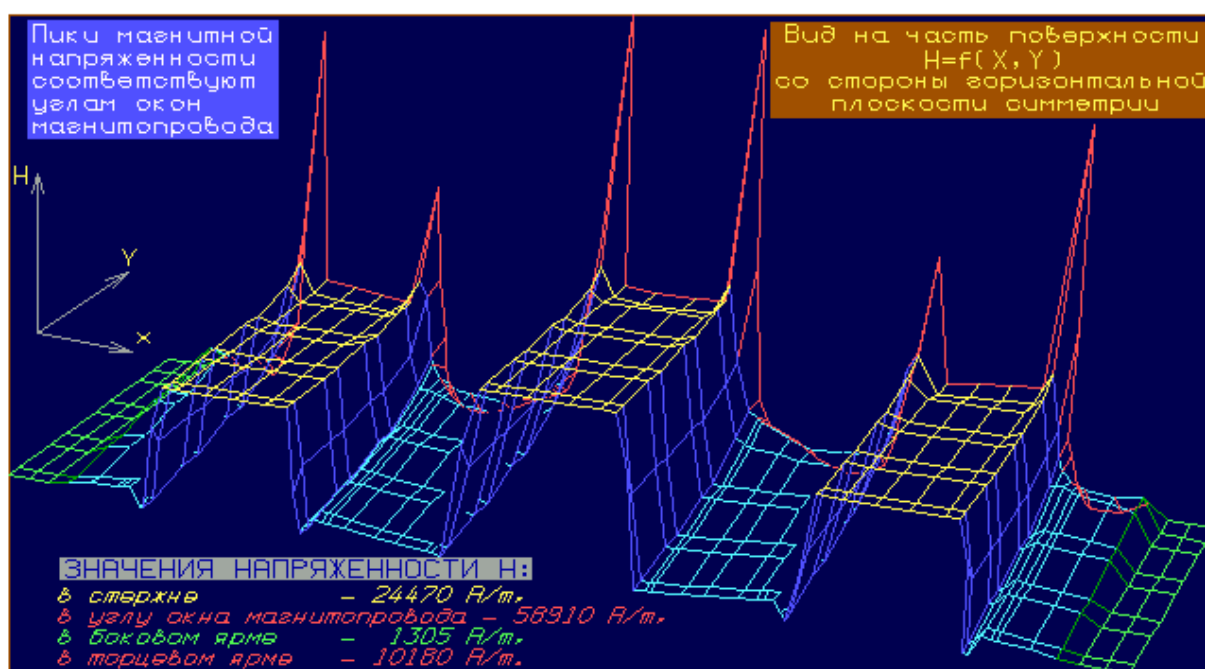
<u>ВВЕДЕНИЕ</u>	4
1. Термины и определения	6
1.1. Общие понятия	6
1.2. Виды трансформаторов	6
1.3. Магнитная система трансформатора	7
1.4. Обмотки трансформатора	8
1.5. Изоляция трансформатора	8
1.6. Отдельные части трансформатора	9
1.7. Детали конструкции и параметры частей трансформатора	10
1.8. Режимы и процессы	11
1.9. Параметры трансформатора	11
1.9.1. Общие параметры	11
1.9.2. Номинальные данные трансформатора	12
2. Буквенные обозначения элементов в электрических схемах	13
3. Обозначение выводов и ответвлений силовых трансформаторов	14
4. Условное обозначение трансформаторов	16
5. Номинальные данные и устройство исследуемого трансформатора	18
6. Работа №1 ТРЕХФАЗНЫЙ ДВУХОБОМОТОЧНЫЙ ТРАНСФОРМАТОР	
7. Работа №2 ПАРАЛЛЕЛЬНАЯ РАБОТА ТРЕХФАЗНЫХ ТРАНСФОРМАТОРОВ	20
8. Работа №3 НЕСИММЕТРИЧНАЯ НАГРУЗКА ТРЕХФАЗНЫХ ТРАНСФОРМАТОРОВ	37
9. Работа №4 ИССЛЕДОВАНИЕ ЯВЛЕНИЙ, ВОЗНИКАЮЩИХ ПРИ НАМАГНИЧИВАНИИ МАГНИТОПРОВОДА ТРАНСФОРМАТОРА	55
ЛИТЕРАТУРА	98
<u>URL-адреса предприятий, производящих электрические машины трансформаторы</u>	98
<u>URL-адрес РАО ЕЭС "Россия"</u>	98
Приложения:	
1. Паскаль-программа "Расчет характеристик холостого хода трансформатора"	99
2. Трансформаторы специального назначения	102
3. Форма отчета по Лабораторной работе (образец оформления)	141
4. Примерная программа дисциплины Электрические машины	154

Рис. 3. Web-страница с учебным пособием

(<http://zei.narod.ru/soderghanie.html>)



а



б

Рис. 4. Картины пространственного распределения: а – индукции магнитного поля;

б – напряженности магнитного поля (кадры слайд-фильма)

(<http://zei.narod.ru/filmr/film.html>)

**Результаты расчета рабочих характеристик I_1 , η , $\cos\varphi$, $s = f(P_2)$
асинхронного двигателя с короткозамкнутым ротором**

Таблица 1. Исходные данные

$P_{2\text{НОМ}} = 15.00$ кВт	$U_{1\text{НОМ}} = 220/380$ В	$I_{1\text{НОМ}} = 28.40$ А
$P_{\text{СТ}} = 0.369$ кВт	$P_{\text{МЕХ}} = 0.117$ кВт	$I_{0a} = 0.520$ А
$I_{0p} = 7.910$ А	$r_1 = 0.355$ Ом	$r_2' = 0.186$ Ом
$c_1 = 1.025$	$a' = 1.051$	$a = 0.364$ Ом
$b' = 0.000$	$b = 1.650$ Ом	$s_{\text{НОМ}} = 0.024$

Таблица 2. Результаты расчета, выполненного на ПК (см. п.п 7 и 7.1 УМК)

#	s	-	0.005	0.010	0.015	0.020	0.025	0.030	0.024
1	R	Ом	39.46	19.91	13.40	10.14	8.18	6.88	8.51
2	X	Ом	1.650	1.650	1.650	1.650	1.650	1.650	1.650
3	Z	Ом	39.50	19.98	13.50	10.27	8.35	7.08	8.67
4	I_2''	А	5.57	11.01	16.30	21.42	26.35	31.09	25.38
5	$\cos\omega_2'$	-	0.999	0.997	0.993	0.987	0.972	0.980	0.982
6	$\sin\omega_2'$	-	0.042	0.083	0.122	0.161	0.198	0.233	0.190
7	I_{1a}	А	6.09	11.49	16.70	21.66	26.35	30.76	25.44
8	I_{1p}	А	8.14	8.82	9.90	11.35	13.12	15.16	12.74
9	I_1	А	10.17	14.49	19.41	24.45	29.44	34.29	28.45
10	I_2'	А	5.71	11.29	16.71	21.95	27.01	31.87	26.02
11	P_1	кВт	4.02	7.59	11.02	14.30	17.39	20.30	16.79
12	$P_{\text{э1}}$	кВт	0.11	0.22	0.40	0.64	0.92	1.25	0.86
13	$P_{\text{э2}}$	кВт	0.02	0.07	0.16	0.27	0.41	0.57	0.38
14	$P_{\text{доб}}$	кВт	0.02	0.04	0.06	0.07	0.09	0.10	0.08
15	ΣP	кВт	0.63	0.82	1.10	1.46	1.90	2.41	1.81
16	P_2	кВт	3.38	6.77	9.92	12.83	15.49	17.89	14.98
17	η	-	0.842	0.892	0.900	0.898	0.891	0.881	0.892
18	$\cos\omega$	-	0.599	0.795	0.860	0.886	0.895	0.897	0.894

Номинальные данные спроектированного двигателя:

$P_{2\text{НОМ}} = 15$ кВт, $U_{1\text{НОМ}} = 220/380$ В, $I_{1\text{НОМ}} = 28.4$ А, $\cos\omega_{\text{НОМ}} = 0.894$, $\eta_{\text{НОМ}} = 0.892$.

Методические рекомендации студенту-расчетчику по анализу результатов:

- 1) сравнить рассчитанное значение мощности P_2 на валу двигателя (см. в табл. 2 строку 16) с заданным значением P_2 (см. Задание);
- 2) сравнить рассчитанные значения кпд η и коэффициента мощности $\cos\omega$ (см. в табл. 2 строки 17 и 18) с принятыми значениями (см. Задание).

Если рассчитанные величины отличаются от указанных **более чем на 15 %**, то необходимо внести коррективы в расчет и выполнить его еще раз.

Рис. 5. Текст файла с результатами расчета [рабочих характеристик асинхронного двигателя](http://zei.narod.ru/Technical_Requirement_AD.pdf) (http://zei.narod.ru/Technical_Requirement_AD.pdf)

(в качестве теста приняты данные, изложенные в книге: Проектирование электрических машин: учеб. пособие для вузов / под ред. И.П. Копылова. – М.: Высшая школа, 2002. – С.467)

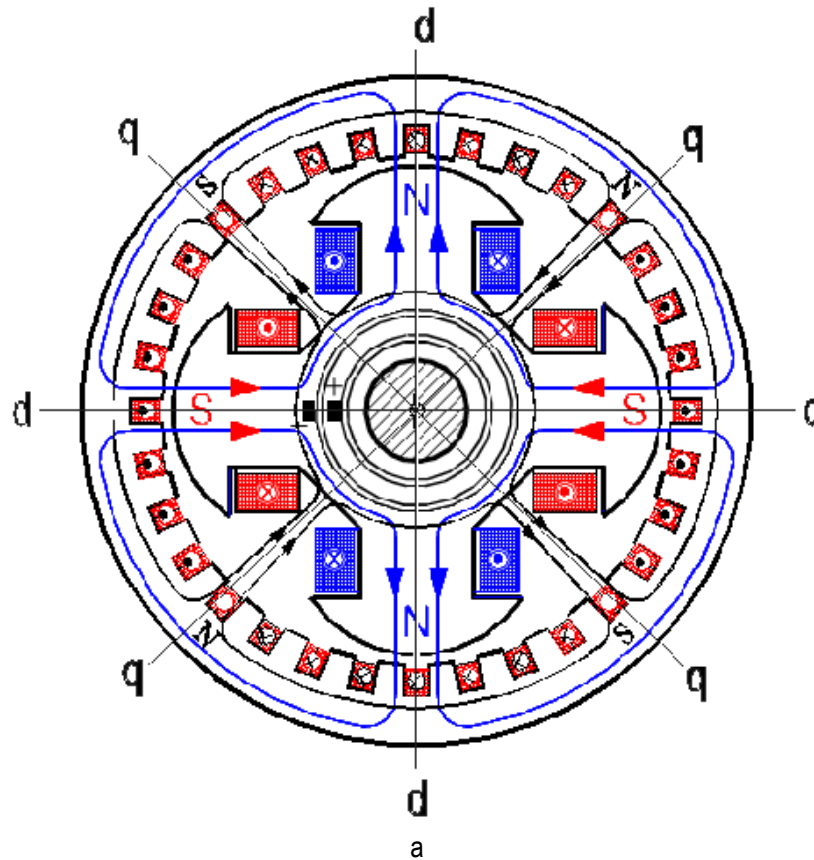


а



б

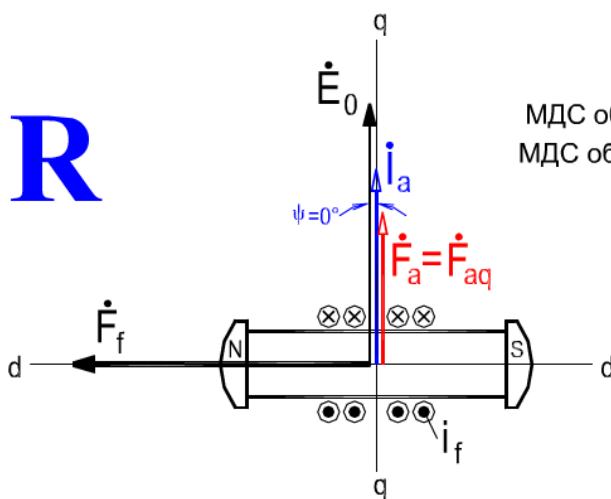
Рис. 6. Скриншоты кадров компьютерного видеofilьма «Конструкция и технология электрических машин». Часть 1. Трансформаторы:
 а – фотография трансформатора, установленного на Асуанской ГЭС;
 б – осевая опрессовка и сушка обмоток НН и ВН трансформатора



$$i_f \rightarrow \dot{F}_f \rightarrow \dot{\Phi}_f \rightarrow \dot{E}_0 \rightarrow i_a \rightarrow \dot{F}_a = \dot{F}_{aq}$$

Диаграмма построена.
Угол $\psi = 0$ градусов.

R



МДС обмотки якоря $F_a = F_{aq}$ направлена по оси **q**.
МДС обмотки возбуждения F_f направлена по оси **d**.
Реакция якоря является поперечной.

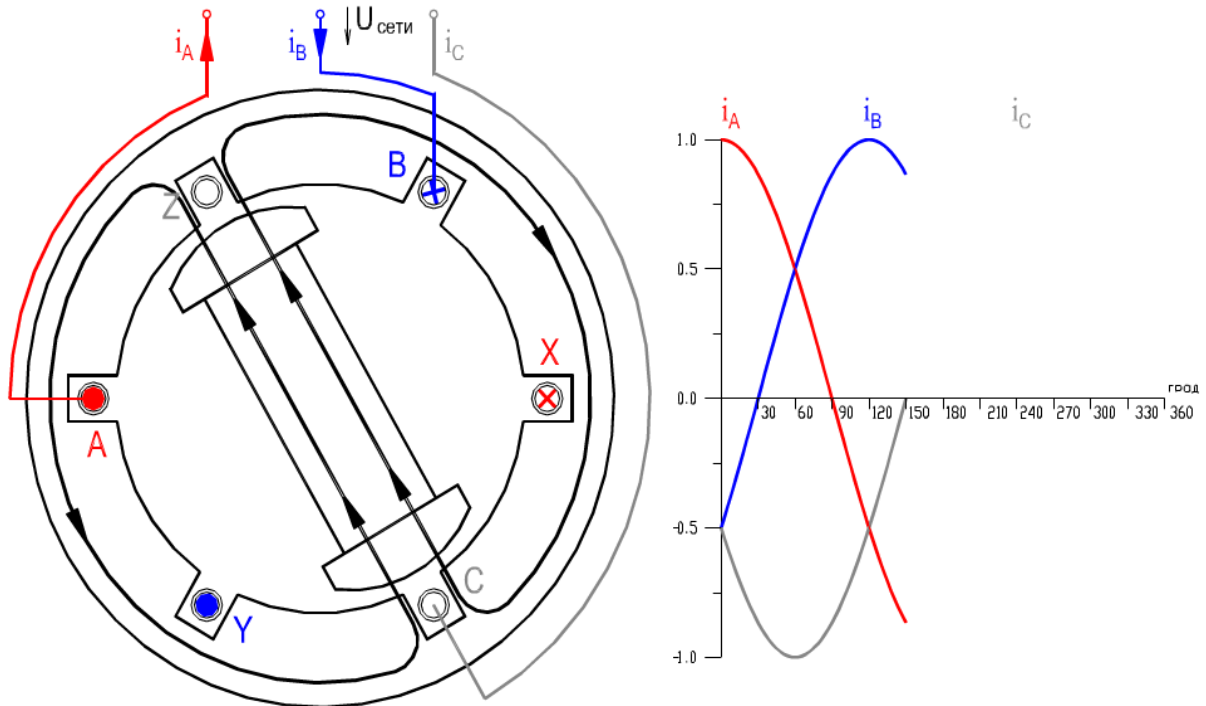
РЕАКЦИЯ ЯКОРЯ ЯВНОПОЛЮСНОГО СГ
ПРИ АКТИВНОЙ НАГРУЗКЕ

б

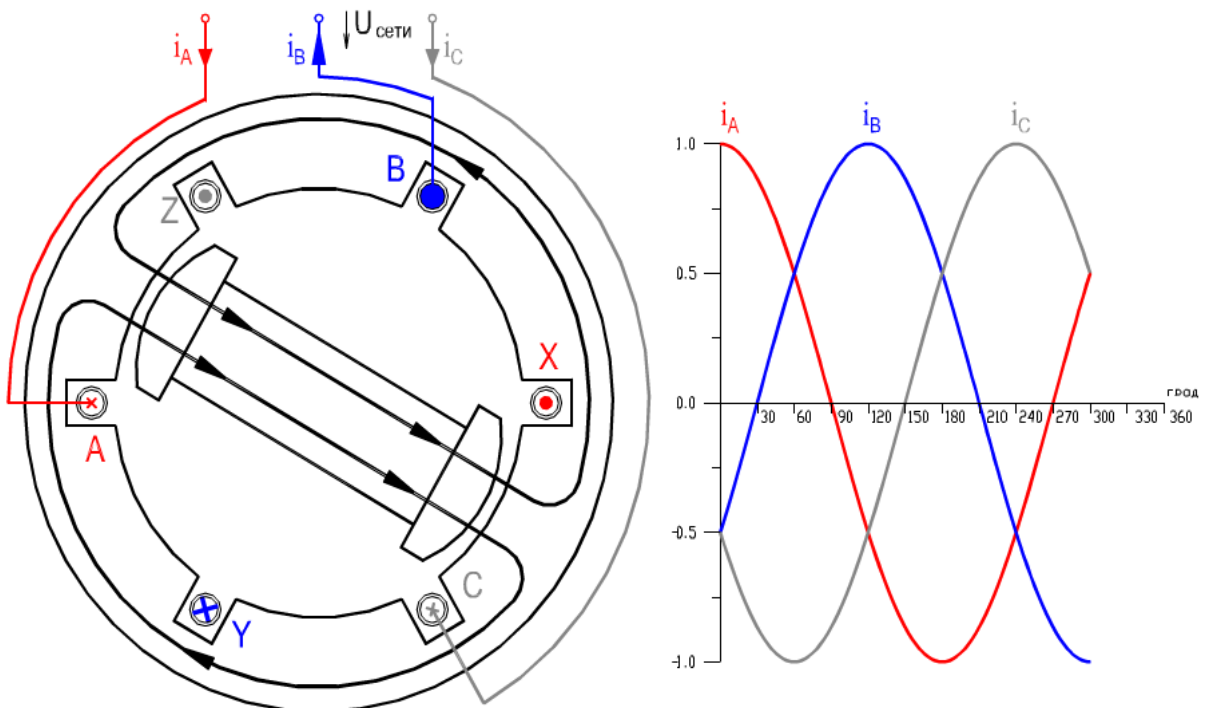
Рис. 7. Мультимедийные фрагменты видеолекции «Реакция якоря синхронной машины», иллюстрирующие в динамике влияние активной нагрузки, подключенной к обмотке якоря: а – взаимная ориентация магнитных полей якоря и индуктора; б – векторная диаграмма

УСЛОВИЯ СОЗДАНИЯ ВРАЩАЮЩЕГОСЯ МАГНИТНОГО ПОЛЯ ТРЕХФАЗНОЙ РАСПРЕДЕЛЕННОЙ ОБМОТКОЙ:

1. Фазы обмотки смещены в пространстве на 120 электрических градусов;
2. Токи в фазах смещены во времени на 120 градусов



а

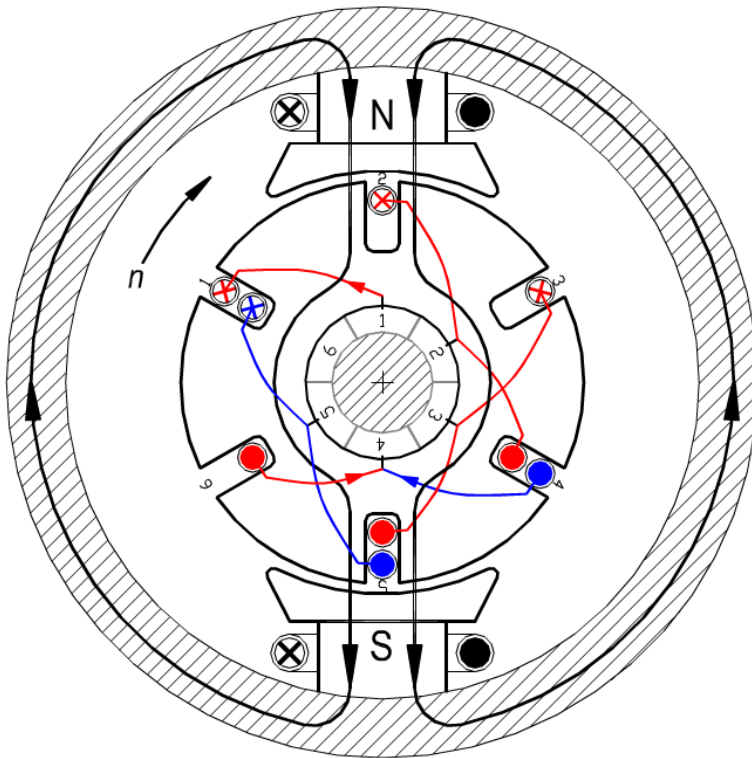


б

Рис. 8. Фрагменты видеоматериалов, иллюстрирующие в динамике условия создания вращающегося поля трехфазной обмоткой, соответствующие различным мгновенным значениям токов в обмотке:

$$a - i_b = |-i_a| \approx 0,707, i_c = 0; \quad б - i_a = i_c = 0,5, i_b = -1,0$$

УКЛАДКА БАРАБАННОЙ ОБМОТКИ В ПАЗЫ ЯКОРЯ МАШИНЫ ПОСТОЯННОГО ТОКА

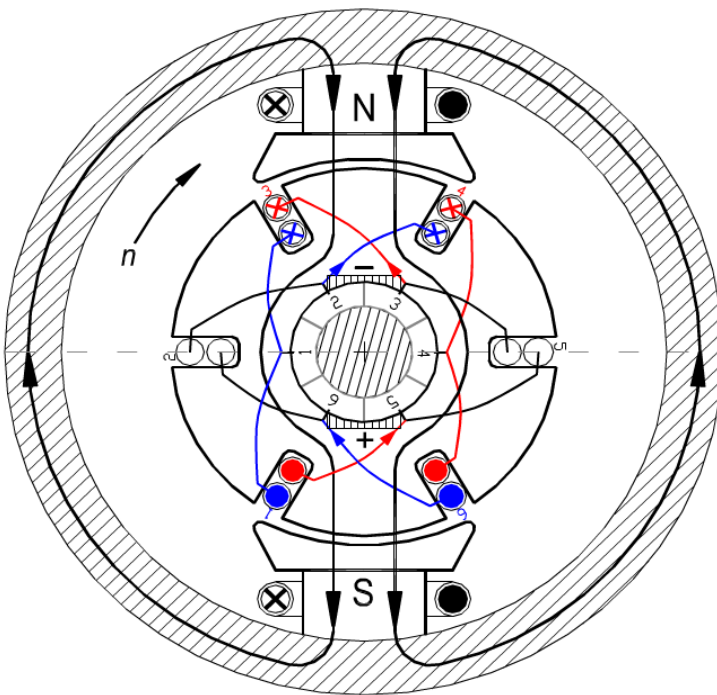


а

	Параллельная ветвь обмотки	
	Первая	Вторая
# СЕКЦИИ	1	4
	2	5
	3	

Укладка 5-й секции.
Начало 5-й секции подсоединяется к 5-й коллекторной пластине, а левая сторона секции укладывается наверх 5-го паза

ПРИ ВРАЩЕНИИ В ОБМОТКЕ ЯКОРЯ ИНДУЦИРУЕТСЯ ПЕРЕМЕННАЯ Э.Д.С.



б

Г.Н.

Положение якоря	## секции в ветви		## к.з. секций
	1-я ветвь	2-я ветвь	
0°	1, 2 и 3	4, 5 и 6	нет
30°	1 и 2	4 и 5	3 и 6
60°	6, 1 и 2	3, 4 и 5	нет
90°	6 и 1	3 и 4	2 и 5
120°	5, 6 и 1	2, 3 и 4	нет
150°	5 и 6	2 и 3	1 и 4
180°	4, 5 и 6	1, 2 и 3	нет
210°	4 и 5	1 и 2	6 и 3
240°	3, 4 и 5	6, 1 и 2	нет
270°	3 и 4	6 и 1	5 и 2

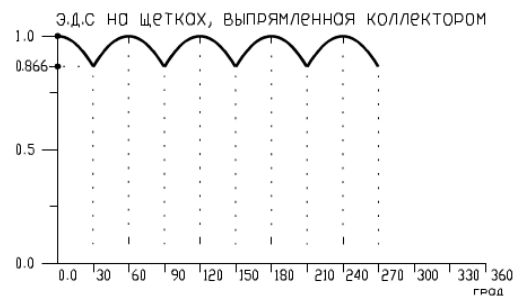


Рис. 9. Фрагменты видеоматериалов, иллюстрирующие в динамике процесс укладки и устройство якорной обмотки машины постоянного тока (а), а также выпрямление напряжения коллектором генератора (б)

Вывод

В результате использования преподавателями и студентами разработанных Учебно-методического комплекса дисциплины «Электрические машины» (электронная реализация УМК размещена в Интернете, <http://zei.narod.ru/ind.html>), и мультимедийных приложений повышена эффективность изложения материала, достигнуто понимание студентами сложных вопросов теории и устройства электрических машин и трансформаторов.

Библиография

1. [Забудский Е.И. Математическое моделирование управляемых электроэнергетических устройств](#): Учебное пособие для вузов. - Ульяновск: УлГТУ, 1998. - 120 с.
2. [Забудский Е.И. Анализ управляемых электроэнергетических устройств методом конечных элементов](#): Учебное пособие для вузов. - М.: МГАУ им. В.П. Горячкина, 1999. - 141 с.
3. [Забудский Е.И. Электрические машины. Ч. 1. Трансформаторы](#): Учебное пособие для вузов.– М.: МГАУ им. В.П. Горячкина, 2002. - 168 с.
4. [Забудский Е.И. Электрические машины. Ч. 3. Синхронные машины](#): Учебное пособие для вузов – М.: МГАУ им. В.П. Горячкина, 2008. - 196 с.
5. [Забудский Е.И. Электрические машины. Ч. 4. Машины постоянного тока](#): Учебное пособие для вузов – М.: МГАУ им. В.П. Горячкина, 2010. - 217 с.
6. [Забудский Е.И. Компьютерный фильм “Геометрическая интерпретация результатов расчета магнитного поля в устройствах электромеханики”](#): Пакет компьютерных программ / Зарегистрирован в Отраслевом фонде алгоритмов и программ НИИВО Минобразования РФ. Регистр. № 94696/16.02.94. [Сертификат №3/93 от 10.09.93г.](#)
7. Evgeny Zabudsky. [Lecturer's web-site and its role in distance learning](#) // International Journal “Information Technologies and Knowledge” (IJ ITK), Vol. 1/2007, p. 279-281.
8. А.с. 1651326 СССР. Трехфазный насыщающийся реактор / Е.И. Забудский, Ю.В. Ермураки, С.Ф. Козырин. Опубл. 23.05.91, Бюл. # 19.
9. А.с. 1658224 СССР. Трехфазный управляемый реактор / Е.И. Забудский, Ю.В. Ермураки, С.Ф. Козырин. Опубл. 23.06.91, Бюл. # 23 (<http://zei.narod.ru/11.html>).
10. А.с. 1663721 СССР. Трехфазный статический ферромагнитный утроитель частоты / Е.И. Забудский, Ю.В. Ермураки, С.Ф. Козырин, В.Г. Ольховский, О.В. Воловецкий. Опубл. 15.07.91, Бюл. # 26.

Информация об авторе



Евгений Забудский – Российский университет Дружбы народов, профессор кафедры Кибернетика и мехатроника, доктор техн. наук, 117923, Москва, Россия, ул. Орджоникидзе, д. 3; e-mail: zei@inbox.ru; web-сайт: <http://zei.narod.ru>.

Направление научной работы: Регулируемые электромагнитные устройства для управления режимами электроэнергетических систем. Компьютерные информационные технологии в образовании.

СОДЕРЖАНИЕ НЕКОТОРЫХ ЭТАПОВ СИСТЕМНОГО АНАЛИЗА И СИНТЕЗА (ПРОДОЛЖЕНИЕ «МЕТОДОЛОГИЧЕСКОГО МЕМОРАНДУМА» – ЧАСТЬ III)

Анатолий Крисилов

Аннотация: Рассмотрены методологические вопросы системного анализа и синтеза, разные «схемы» анализа. Дана характеристика видов неопределенности и пути ее преодоления в работах участников КДС; Показаны связи инженерии знаний и когнитологии.

Ключевые слова: системный анализ и синтез, самоопределение, виды синтеза, неопределенность, механизмы развития, инфос, понимание, моделирование, инженерия знаний, когнитология.

Summary: The methodological approach to obtaining description of the system analysis and synthesis is used. A good deal of attention is paid to the uncertainty and its decreasing in works of KDS-participants. Some links between the knowledge engineering and cognitology are shown.

Keywords: system analysis and synthesis, synthesis' kinds, uncertainty, adaptation, development' mechanisms.

ACM Classification Keywords: A.1. Information Technology for Knowledge Management. H.1.1 – Systems and Information Theory – General Systems Theory.

Не в напрасном тщании убедить читателя доказательствами, но в неуклонном внутреннем стремлении автора обрести и явно выразить смутно ощущаемое им единство мира написан этот трактат.

А. Подводный. «Тонкие тела», ч. I

Небесные тенета имеют крупные ячейки. Но ничто не может миновать их.

Лао-цзы. «Дао де цзин»

1. Вступительные замечания

Не поклажа убивает животное, а избыток ее.

М. Сервантес

На протяжении последних двух лет в рамках KDS (конференции, сборники) было опубликовано несколько материалов, объединенных названием «Методологический меморандум», части I и II [1, 2], а также работа [3]. Кроме того, в развитие и продолжение этих работ появились публикации [4, 5]. Настоящий материал продолжает поднятую ранее тему, – все более прочным становится представление, что продуктивное движение в области проблематики KDS, в области работ по искусственному интеллекту, по инженерии знаний, по решению ноосферных задач требует сейчас все более глубокого применения серьезного методологического инструментария, в частности, – аппарата системологии и смежных областей знания.

В предлагаемой работе, поневоле конспективной, акцент сделан на следующих вопросах. Продолжается рассмотрение содержания, целей и особенностей основных этапов системного анализа и системного синтеза, – с использованием этого аппарата для формирования Интеллектуального Агента [3]. В свою очередь, эту структуру предполагается использовать при построении систем искусственного интеллекта, для моделирования сложных систем, как инструмент описания и формирования знаний, наконец, как инструмент обучения. Здесь следовало бы обратить внимание на такой важный аспект указанной работы: одна из ориентаций систем искусственного интеллекта, по нашему мнению, – исследование с его помощью интеллекта естественного.

Необходимо подчеркнуть, что, занимаясь системным анализом сложных качественных объектов, планируя и выстраивая этапы системного синтеза сложных (многоцелевых, многоуровневых, гетерогенных, полиморфных и т. д.) структур, мы обязательно должны совершать последовательные шаги познания, выявлять и конструировать смыслы, сознательно расширять границы познанного. Первым инструментом, который у нас имеется для этих целей, является наше собственное сознание. В нем формируется более или менее адекватный образ объективной реальности, той живой среды, с которой мы имеем дело в тех или иных задачах и которая составляет ткань нашей деятельности и жизнедеятельности в целом. В этой среде, в объективной реальности работают наши воспринимающие механизмы, перцептивные системы, – они создают основу, отправляясь от которой можно двигаться в глубь вещей, в чем, собственно, и состоит сущность мышления и его прямое назначение.

Сказанное, в частности, поясняет, почему в целом ряде аспектов предметом нашего обсуждения будут вопросы мыслительной деятельности, гносеологии, новой науки – когнитологии. Если мы хотим, чтобы результаты системного анализа (и, естественно, синтеза) были адекватны поставленным задачам, целям, требованиям, – нам не избежать таких выходов к возможно более широким постановкам и представлениям. Мало того, каждый разработчик, повидимому, должен быть готов к сознательной ревизии своих прежних представлений, к формированию новых разнообразных фильтров, к сознательной и усиленной эксплуатации своего сознания. (В этом последнем обороте нет тавтологии, содержательно именно так обстоит дело, и от сознательного включения определенных возможностей и форм сознания выигрывает, в первую очередь, сам включающий).

В целом нужно сказать, что в настоящей работе ставится, вообще говоря, практическая, даже прагматическая задача: а) рассмотреть основные этапы (их содержание и особенности) системного анализа и системного синтеза как инструментов работы со сложными объектами и процессами, имеющими в основном качественную природу, слабоформализованными и плохо поддающимися структуризации; б) построить, в частности, при помощи этих инструментов, некоторую информационную структуру, такого Интеллектуального Агента, который был бы в состоянии эффективно функционировать в сложной информационной среде. При этом параллельно, разумеется, необходимо будет затронуть некоторые смежные вопросы и проблемы.

Вопрос об упомянутых выше фильтрах, через которые (и при помощи которых!) мы воспринимаем и/или перерабатываем в нашем сознании все воспринятое и «придуманное», чрезвычайно важен для реализации шагов системного анализа и синтеза. Мы, безусловно, обратимся к нему, но несколько позднее. Сперва – несколько других общих вопросов.

2. Различные «схемы» системного анализа и некоторые определения

Язык дан человеку, чтобы скрывать свои мысли.

Околонаучный фольклор

За последние десятилетия накопилось весьма большое количество работ по системному анализу и весьма небольшое – по системному синтезу. В целом ряде публикаций на разные темы встречается ссылка на применение модного метода - системного анализа; достаточно некому автору уразуметь, что он имеет дело со сложным объектом, как все, что он с ним делает, он считает системным анализом. В то же время в серьезной литературе давно появились относительно устойчивые представления об этом предмете. Приведем для примера и для ясности несколько определений (цит. по [12]).

«Энциклопедия кибернетики» (Киев, 1974 г., Гл. ред. Укр. Сов. Энцикл.): «...методология исследования объектов путем представления их в качестве *систем* и анализа этих систем. Является эффективным средством решения сложных, недостаточно четко сформулированных проблем. Предмет системного анализа включает в себя общесистемные характеристики и [описание] взаимодействия системы с ее окружением. Важнейшими теоретическими сферами системного анализа являются следующие направления: общие принципы проведения исследований сложных систем (и принципы интегрирования различных методов); проблемы сложности, неопределенности и методы их разрешения, проблема предельных характеристик системы; принципы машинно-реализуемой имитации. Ведущей концепцией системного анализа служит системный подход».

«Советский энциклопедический словарь» (М., 1989 г., «Сов.Энциклопедия»): «...совокупность методов и средств, используемых для подготовки и обоснования решений по сложным проблемам политического, военного, социального, экономического, научного и технического характера. Основная процедура - построение обобщенной модели, отображающей взаимосвязи реальной ситуации. Опирается на системный подход».

«Философский словарь» (М., 1986): «...согласно принципам системного анализа, возникающая перед обществом та или иная сложная проблема (прежде всего, проблема управления) должна быть рассмотрена как нечто целое, как система во взаимодействии всех ее компонентов. Для принятия решения об управлении этой системой необходимо определить ее цель. Важным этапом системного анализа является построение обобщенной модели исследуемой или конструируемой системы»... И т. д.

Таблица I. Разновидности поэтапных структур системного анализа

По С. Оптнеру	По С. Юнгу	По С. Никанорову	По Ю. Черняку
1. Идентификация симптомов	1. Определение целей организации	1. Выявление проблемы	1. Анализ проблемы
2. Определение актуальности проблемы	2. Выявление проблемы	2. Оценка актуальности проблемы	2. Определение системы
3. Определение целей	3. Диагноз	3. Анализ ограничений проблемы	3. Анализ структуры системы
4. Определение структуры	4. Поиск решения	4. Определение критериев	4. Формулировка общих целей и критериев
5. Определение возможностей	5. Оценка и выбор альтернативы	5. Анализ существующей системы	5. Декомпозиция целей, определение потребностей в ресурсах, композиция целей
6. Выявление альтернатив	6. Согласование решения	6. Поиск возможностей	6. Выявление ресурсов, композиция целей
7. Оценка альтернатив	7. Утверждение решения	7. Выбор альтернативы	7. Прогноз и анализ будущих условий
8. Выработка решения	8. Подготовка к вводу в действие	8. Обеспечение признания альтернативы	8. Оценка целей и средств
9. Верификация и принятие решения	9. Управление реализацией решения	9. Принятие решения (признание формальной ответственности)	9. Выбор вариантов
10. Запуск процесса реализации решения	10. Оценка эффективности	10. Реализация решения	10. Диагноз существующей системы
11. Управление процессом реализации решения		11. Определение результатов решения	11. Разработка комплексной программы развития
12. Оценка реализации и ее последствий			12. Проектирование организации для достижения целей

Из приведенных текстов видны основные черты системного анализа, – методологии, разрабатываемой достаточно давно (по меркам существующих темпов научно-технического развития), однако все еще находящейся на стадии становления. Предлагаемая нами на KDS серия методологических работ является иллюстрацией высказанной оценки.

Выше на **Табл. I** приведена в кратком изображении сводная картина нескольких отличающихся «схем» проведения системного анализа. Они принадлежат различным известным школам и разработчикам, в них много общего: привязка к работам и задачам по управлению производством или крупными научно-техническими проектами, опора на мощный административно-бюрократический ресурс, идеология субъектно-объектных отношений (!). Различия между ними невелики, относятся больше к технико-технологическим вопросам и к разной зависимости проектантов, разработчиков и работников от степени бюрократизации.

Поскольку эта сводная таблица находится перед глазами, целесообразно для сопоставления практически без комментариев привести опубликованный ранее, в «Кратком методологическом меморандуме» [1] перечень предлагаемых нами этапов системного анализа:

1. Самоопределение.
2. Целеполагание; постановка задачи.
3. Первичная структуризация.
4. Декомпозиция или вторичная структуризация.
5. Выявление основных принципов и закономерностей в данной системе.
6. Основные качественные и количественные характеристики данной системы.
7. Анализ основных процессов в данной системе, имеющих в ней механизмов.
8. Анализ ресурсов.
9. Моделирование.
10. Анализ результатов; формулировка новых целей.

Следует отметить, что предложенная структура отличается не только наличием очень важных этапов (таковым, например, является этап самоопределения, а также моделирование, по непонятным причинам отсутствующее в вышеприведенных вариантах), но и свойственным именно современному системологическому подходу целостным, холистическим мышлением, отношением к системе связей в исследуемой системе, что и нашло частичное отражение в предыдущих частях данной работы.

Рассмотрим далее вопрос, относящийся к содержанию (и ограничениям) в работе человека, имеющего дело со сложными процессами и объектами: это неопределенность.

3. Неопределенность, адекватность отражения, полнота модели

Что мы знаем о лисе?

Ничего.

И то – не все.

Б. Заходер

Совершенно естественно, что в этой задаче (описание, оценка, анализ, наконец, - пути преодоления) неопределенности сошлись интересы многих разработчиков в области инженерии знаний. Это имеет место и для активных участников движения KDS. Попробуем вкратце описать эту ситуацию, выявляя достаточно широкие совместные возможности осмыслительного и аппаратного характера.

Известна следующая классификация нечеткости/неопределенности в различных задачах работы со знаниями (мы сознательно здесь не обсуждаем вопрос о корректности или некорректности применения в данных случаях использованных терминов, для нашей задачи это не очень важно):

- а) многозначность, многоаспектность рассматриваемых данных или знаний;
- б) ненадежность данных, их случайный характер;
- в) недетерминированность исходных данных или выводов;
- г) неполнота рассматриваемой картины или описываемого процесса;
- д) нечеткость или неточность описания.

Другим способом классификации может послужить фиксация места или источника неопределенности - на выходе измерительного устройства, в канале передачи данных, на выходе компьютера и т. д.

В работах В. С. Донченко [6 и некот. др.] рассмотрены математические модели, определенным образом специфицированные для работы с различными видами неопределенностей: мы видим здесь множественные модели (МнМоН), работающие со случайностями, с нечеткостью, с недетерминированностью. Рассматривая (условную и содержательную) структуру неопределенности, автор предлагает для разных компонент этой структуры разные подходы для уменьшения неопределенности, например, минимаксные схемы или интервальный подход, и для вариативных и функциональных частей рассматриваемых величин оказывается возможным в разной степени такое уменьшение.

Решая (в других работах, вместе с Н. Ф. Кириченко) задачу кластеризации по гиперплоскостям, авторы используют «расстояние Махаланобиса», вводят понятие «расстояние соответствия». Это представляется чрезвычайно интересным и могущим быть эффективно использованным в работе Г. С. Воронкова [8], посвященной ревизии ряда серьезных положений теории отражения. В этой работе автор предлагает даже, отмечая не обязательно изоморфный характер отображения информации в нейронных структурах мозга, вместо названия «теория отражения» использовать термин «теория соответствия». Возможно, теперь, при использовании понятия о расстоянии соответствия, можно навести определенную метрику, отражающую степень сходства воспринятого и воспринимаемого образов. (В отношении названия: может быть, можно даже сказать «теория ответственности», акцентируя внимание не только и не столько на результате, финальном акте, - «отражении», - но и на делящемся динамическом процессе, и на состоянии «картины отражения», впрочем, это не обязательная тонкость).

Хотелось бы указать еще на две-три связи в рассматриваемой непростой проблеме. Одна из них - фиксация принципиальной возможности-невозможности и адекватности моделирования сложных (физических, но не только!) процессов, схематически изображенная на одном из коронных рисунков (рис. 10.1.) главы 10 в работе И. И. Горбаня [7], представленная на одном из докладов автора в 2009 году. Представляется, что и здесь может иметь смысл попытка оценки адекватности при помощи «расстояния соответствия». Кажется, что возникают любопытные связи, и не только формальные.

Далее, инженерным средством для последовательного уточнения описания сложных объектов и ситуаций, в том числе, и с использованием указанных здесь идей, являются растущие сети [9 и др.]. Если рассматривать их не только как пирамидальные (то есть именно как сетевые, не обязательно иерархические) и, кроме того, обеспечить в возникающих узлах и ветвях (при определенных условиях) некоторую свободу действий, то их информационные возможности могут быть ощутимо расширены.

Наконец, следует сказать о мощной роли идеи «инфоса» (см., напр., [15] и ряд др. работ тех же авторов). Представляется, что Инфос может играть роль (в том числе, в приведенных выше задачах) отличного буфера, контрольного или метрологического агента, передатчика и/или преобразователя информации, линии задержки, концентратора, фильтра и так далее. Он вправе требовать больше внимания сегодня.

4. Вместо заключения

Большие батальоны всегда правы.

Наполеон

Настоящий материал представляет собой третий фрагмент начатого два года назад «Методологического меморандума». Хотя собственно методологическая/системологическая часть этой работы далеко не исчерпана (и ее участники надеются на реальное и основательное ее продолжение), тем не менее, в процессе работы оказались затронутыми и, в некоторой степени, раскрытыми и определенные содержательные аспекты работы со знаниями. Это относится к разработке Интеллектуального Агента, к описанию эффективных методов системного синтеза (реистического, атрибутивного, реляционного), к постановке и разработке этапа самоопределения и др.

Говорят, что если ты хочешь рассмешить Бога, – расскажи ему о своих планах. Все же стоит рискнуть. За пределами настоящей работы осталось в данный момент несколько важных вопросов, которые просятся к рассмотрению в ближайшее время. К ним относятся:

- - раскрытие структуры Интеллектуального Агента; описание системы «Вольвокс»;
- - описание роли положительной и отрицательной обратной связи в системных объектах;

- - включение эмерджентных, непредусмотренных ситуаций («Штирлиц идет по коридору...»);
- - описание разных совокупностей фильтров;
- - детальное рассмотрение видов системного синтеза; их комбинирование в конкретных задачах;
- - решение в системологическом аспекте задач таксономии и мерономии;
- - рассмотрение принципа Кюри (в системном представлении);
- - роль вакантных позиций;
- - роль понятий «граница», «застава»;
- - формы изменений и механизмы развития; виды адаптации,

и еще некоторые.

В заключение нашего разговора о желаемой (возможной ли?) полноте и точности выстраиваемого описания сложных процессов следует привести несколько цитат.

Ю. М. Лотман: «...Возникает центральный вопрос и вытекающие из него следствия: 1) неизбежность того, что пространство реальности не охватывается ни одним языком в отдельности, только их совокупностью; 2) необходимость более чем одного (минимум двух) языков для отражения запредельной реальности».

Нильс Бор, автор принципа дополнительности, устанавливающего несводимость и взаимную (целостную) дополнительность дискретного и волнового (средового!) языков описания очень тонких фундаментальных событий: «...урок, который мы из этого кризиса извлекли, решительно продвинул нас по пути никогда не кончающейся борьбы за гармонию между содержанием и формой. Описать процессы, протекающие в окружающем нас мире, с помощью одного языка невозможно. Необходимы разные описания, в каждом из которых яснее проявляются те или иные (взаимно-исключающие и дополняющие друг друга) особенности изучаемого явления». Для некоторых задач управления и бизнеса эти вопросы рассмотрены в [13, 14].

В. И. Вернадский: «...Рост научного знания XX века быстро стирает грани между отдельными науками. Мы все больше специализируемся не по наукам, а по проблемам».

К сожалению, уже нет сейчас среди живых замечательного человека и исследователя Валентина Васильевича Налимова. В контексте представлений о самосознании, самоопределении и развитии в целом очень примечательной и характерной для его образа мышления является концовка одной из последних его работ [10]: «...Теория самоорганизации, если еще и не может достаточно внятно объяснить такое величественное явление, как спонтанное саморазвитие Универсума в целом (включая и развитие сознания в нем), то по крайней мере не должна препятствовать возможности такого явления. Мне кажется, мировоззрение, выступающее перед нами вне самостоятельного семантического начала, недостаточно для осмысления мироустройства. Слишком простым и скучным был бы Мир, лишенный свободных смыслов». И далее: «Сейчас у нас появляется все больше и больше возможностей для обоснования тезиса о вездесущности хотя бы слабых форм сознания. Вселенная предстает перед нами как самосознающая грандиозная структура. Сознание по своей природе спонтанно, а спонтанность порождает самоорганизацию. Самоорганизация - это выбор, или - иначе - фильтрация из потенциально заложенного изначально. Процесс фильтрации можно описать математически - силлогизмом Байеса. Мы можем раскрыть содержательность этой модели. Особенно хорошо это получается для высоких форм сознания - психики человека. И больше ничего сказать нельзя. Дальше - запредельное.

Тайна».

Здесь сохранено все так, как он написал, включая расположение текста. Можно лишь добавить к этому сейчас, по крайней мере, два соображения, два «понимания»:

- постоянно происходит изменение расположения и формы этой границы, она постоянно сдвигается (но, разумеется, не параллельно самой себе), – таков, повидимому, один из законов развития ноосферы;
- другой его, развития, закон – тайна постоянно и неукоснительно остается, и тоже, повидимому, меняется, в том числе – и благодаря нашим изменениям.

Благодарности

Настоящая работа была выполнена при поддержке интернационального проекта ITHEA XXI Института информационных теорий и их приложений FOI ITHEA и Ассоциации ADUIS Украина (Ассоциация разработчиков и пользователей интеллектуальных систем).

Литература

- А. Крисилов, Е. Соловьева, А. Уемов. Краткий методологический меморандум – ч. I. Information Science & Computing. International Book Series: Knowledge – Dialogue – Solution, N. 15. – ITHEA, Sofia, 2009.
- А. Крисилов, А. Уемов. О шагах системного синтеза (продолжение «Методологического меморандума» – часть II). Information Models of Knowledge, v. 19. Proc. of Intern. Confer. “Knowledge – Dialogue – Solution” (KDS – 2010), Kiev, Sept. of 2010, ITHEA, Kiev – Sofia, 2010.
- А. Крисилов. «Шаги системного синтеза и построение Интеллектуального Агента в задачах представления знаний», Труды Междун. Конфер. «КДС - 2010». Варна, июнь 2010.
- Крисилов А. Системный анализ с точки зрения системологии. //«Науковий вісник Міжнародного Гуманітарного Університету», № 1. – Одеса: МГУ, 2011.
- Крисилов А. Краткое описание шагов системного синтеза. //«Науковий вісник Міжнародного Гуманітарного Університету», № 2. – Одеса: МГУ, 2011.
- В. С. Донченко. Структуры, неопределенность: математическое моделирование. //Intern. Book Series “Artificial Intelligence and Decision Making”. № 7, Sofia, 2008.
- И. И. Горбань. Теория гиперслучайных явлений. - Киев: ГП «УкрНИУЦ», 2007.
- Г. С. Воронков. Обязательно ли ощущения являются изоморфными «образами» мира: анализ некоторых аспектов теории отражения. //Intern. Book Series “Artificial Intelligence and Decision Making”. № 7, Sofia, 2008.
- V. Gladun, V. Velychko, L. Svyatogor. Hierarchial Three-level Ontology for Text Processing. //Intern. Book Series “Artificial Intelligence and Decision Making”. № 7, Sofia, 2008.
- В. В. Налимов. Самоорганизация как творческий процесс. В кн. «Синергетическая парадигма». – М., Прогресс-Традиция. 2002.
- И. И. Шафрановский. Симметрия в природе. – М.: Недра, 1968.
- М. Ф. Бондаренко, Е. А. Соловьева, С. И. Маторин. Основы системологии. – Харьков: ХТУРЭ, 1998.
- А. Д. Крисилов. Модельное описание процессов развития: механизмы, структура, система целей, индикаторы. Decision Making and Business Intelligence Strategies and Techniques (International Book Series “Information Sciences and Computing”, N 3), Supplement to International Journal “Information Technologies and Knowledge”, v. 2/2008.
- Бондаренко М. Ф., Соловьева Е. А., Маторин С. И., Ельчанинов Д. Б. Системологическая технология моделирования информационных и организационных систем. – Харьков: ХНУРЭ, 2005. б) Е. А. Соловьева. «Системологическое моделирование бизнеса для разработки информационных систем». Proc. XV International Conference “Knowledge – Dialogue – Solution”, vol. # 1, June 20 – 25, 2009, Varna (Bulgaria) – ITHEA, Sofia, 2009.
- Kr. Markov, Kr. Ivanova, I. Mitov. a) Basic Structure of the General Information Theory. IJ ITA, vol. 14, No 1. – ITHEA, Sofia, 2006. b) General Information Theory Basic Formulations). – FOI COMMERCE, Sofia, 2003. c) Крассимир Марков: http://www.foibg.com/ibs_isc/ibs-13/ibs-13-p01.pdf

Информация об авторах

Анатолий Крисилов – Одесская государственная Академия холода, к. т. н., доц. кафедры информационных и коммуникационных технологий; ул. Дворянская, 1/3, Одесса - 82, 65082, Украина; тел. (0482)-632-598; моб. (38097)-291-33-24; E – m: adkrissilov@list.ru